

(11)特許出願公開番号
特開2003-150419
(P2003-150419A)

(43)公開日 平成15年5月23日(2003.5.23)

| (51)Int.Cl. ⁷ | 識別記号 | F I | コード | コード* | コード* |
|--------------------------|-------|---------------|---------|-----------|------|
| G 0 6 F 12/00 | 5 1 4 | G 0 6 F 12/00 | 5 1 4 M | 5 B 0 0 5 | |
| | 5 1 3 | | 5 1 3 D | 5 B 0 7 5 | |
| 12/08 | 5 0 5 | 12/08 | 5 0 5 Z | 5 B 0 8 2 | |
| 17/30 | 1 3 0 | 17/30 | 1 3 0 Z | | |
| | 1 5 0 | | 1 5 0 A | | |

審査請求 未請求 請求項の数18 OL (全 46 頁)

(21)出願番号 特願2001-348168(P2001-348168)

(22)出願日 平成13年11月14日(2001. 11. 14)

(71)出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地

(72)発明者 茂木 和彦
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内

(72)発明者 大枝 高
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内

(74)代理人 100075096
弁理士 作田 康夫

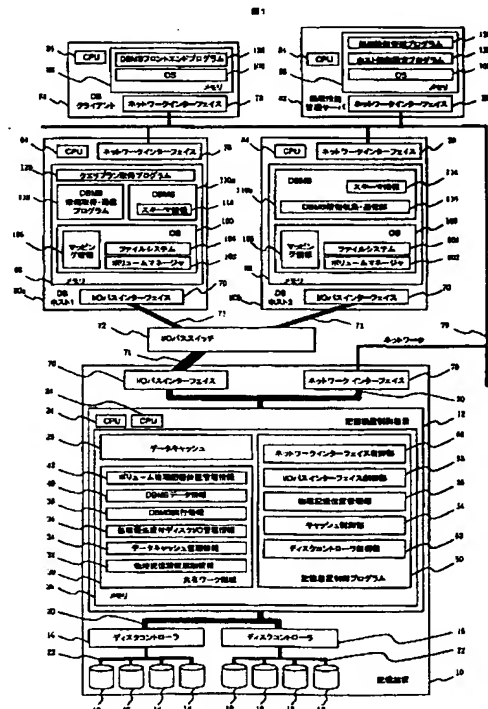
最終頁に続く

(54) 【発明の名称】 データベース管理システムの実行情報を取得する手段を有する記憶装置

(57) 【要約】

【課題】記憶装置においてデータベース管理システム（DBMS）の実行情報やデータベースの処理優先度を加味した制御を行うことにより、好適なデータアクセス性能を実現する。

【解決手段】記憶装置は、DBMSの静的な構成情報をDBMS情報取得・通信プログラム、DBMS情報通信部、ホスト情報設定プログラムを通して取得し、DBMSの実行情報をクエリプラン取得プログラム、DBMS情報通信部、処理性能管理プログラムを通して取得し、処理性能管理プログラムから与えられるデータベースの処理の優先度に関する情報を取得し、これらを処理優先度付ディスクI/O管理情報、DBMS実行情報、DBMSデータ情報中に記憶する。記憶装置制御プログラム内のキャッシュ制御部はこれらの情報を参照してデータキャッシュの制御を行う。



【特許請求の範囲】

【請求項1】データベース管理システムが稼動している計算機との接続手段を有し、

前記データベース管理システムにおけるスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベース管理システムが管理するデータベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報と、前記データベース管理システムが実行する問い合わせ処理の処理実行計画を含む前記データベース管理システムにより管理されるデータベースに関する情報を取得する情報取得手段を有することを特徴とする記憶装置。

【請求項2】前記接続手段を用いて複数の前記データベース管理システムが稼動している計算機と接続することを特徴とする請求項1に記載の記憶装置。

【請求項3】前記情報取得手段が複数の前記データベース管理システムが管理するデータベースに関する情報を取得することを特徴とする請求項1に記載の記憶装置。

【請求項4】前記情報取得手段は前記接続手段を用いて情報を取得することを特徴とする請求項1に記載の記憶装置。

【請求項5】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムから取得することを特徴とする請求項1に記載の記憶装置。

【請求項6】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムとは異なる少なくとも1つのプログラムを通して取得することを特徴とする請求項1に記載の記憶装置。

【請求項7】前記記憶装置は少なくとも1つ以上のデータを記憶する物理記憶手段を有し、前記記憶装置内の前記物理記憶手段に対するアクセス順を決定する物理記憶制御手段を有し、前記物理記憶制御手段が前記物理記憶手段に対するアクセス順を決定する際に前記情報取得手段により取得した情報を利用することを特徴とする請求項1に記載の記憶装置。

【請求項8】キャッシュメモリを有し、前記キャッシュメモリの管理を行うキャッシュメモリ制御手段を有し、前記情報取得手段により取得した情報から前記物理記憶手段に対してこれから行われるアクセス先とアクセス方法を予測するアクセス予測手段を有することを特徴とする請求項7に記載の記憶装置。

【請求項9】前記キャッシュメモリ制御手段において、前記アクセス予測手段を利用したキャッシュの破棄データ選択制御を実施することを特徴とする請求項8に記載の記憶装置。

【請求項10】前記キャッシュメモリ制御手段において、前記アクセス予測手段を利用したデータプリフェッチ実行制御を実施することを特徴とする請求項8に記載の記憶装置。

【請求項11】前記情報取得手段が取得する情報として、前記データベース管理システムが前記スキーマにより定義される同一のデータ構造に属する前記データベースデータをアクセスする際の並列度に関する情報を含むことを特徴とする請求項10に記載の記憶装置。

【請求項12】前記データベースデータの中身を理解するデータベースデータ解釈手段を有し、前記アクセス予測手段において前記データベースデータ解釈手段を利用することを特徴とする請求項10に記載の記憶装置。

【請求項13】前記情報取得手段が取得する情報に、前記データベース管理システムにより管理されるデータベースに与えられた、もしくは実行される処理毎に与えられた処理の優先度に関する処理優先度情報を含むことを特徴とする請求項7に記載の記憶装置。

【請求項14】前記物理記憶制御手段が前記物理記憶手段に対するアクセス順を決定する際に、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータへのアクセスを優先的に実施する制御を行うことを特徴とする請求項13に記載の記憶装置。

【請求項15】キャッシュメモリを有し、前記情報取得手段により取得した情報から前記物理記憶手段に対してこれから行われるアクセス先とアクセス方法を予測するアクセス予測手段を有し、前記アクセス予測手段を利用したデータプリフェッチを実行し、前記処理優先度情報を加味したキャッシュメモリ制御を実施するキャッシュメモリ制御手段を有することを特徴とする請求項13に記載の記憶装置。

【請求項16】前記キャッシュメモリ制御手段において、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータに対して優先的にプリフェッチを実施することを特徴とする請求項15に記載の記憶装置。

【請求項17】前記キャッシュメモリ制御手段において、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータに対してキャッシュメモリ利用量を多く割り当てる制御を実施することを特徴とする請求項15に記載の記憶装置。

【請求項18】前記記憶装置内の物理記憶手段の稼動情報を取得する物理記憶稼動情報取得手段を有し、前記キャッシュ制御手段において、データプリフェッチを実行する際に前記物理記憶稼動情報取得手段により取

得した情報を利用した制御を実施することを特徴とする請求項15に記載の記憶装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データベース管理システムに関する。

【0002】

【従来の技術】現在、データベース（DB）を基盤とする多くのアプリケーションが存在し、DBに関する一連の処理・管理を行うソフトウェアであるデータベース管理システム（DBMS）は極めて重要なものとなっている。特に、DBMSの処理性能はDBを利用するアプリケーションの性能も決定するため、DBMSの処理性能の向上は極めて重要である。

【0003】DBの特徴の1つは、多大な量のデータを扱うことである。そのため、DBMSの実行環境の多くにおいては、DBMSが実行される計算機に対して大容量の記憶装置を接続し、記憶装置上にDBのデータを記憶する。そのため、DBに関する処理を行う場合に、記憶装置に対してアクセスが発生し、記憶装置におけるデータアクセス性能がDBMSの性能を大きく左右する。そのため、DBMSが稼動するシステムにおいて、記憶装置におけるアクセスの最適化が極めて重要である。

【0004】米国特許5317727（文献1）においては、無駄なアクセスの削減や必要なデータの先読みによりDBMSの性能を向上させる技術について開示している。ユーザからの問い合わせ処理（クエリ）を実行する部分において、問い合わせの実行プランやデータアクセス特性、キャッシュメモリ量、I/O負荷等を考慮して、プリフェッチ実行やその量の決定、キャッシュ（バッファ）管理等を行うことによりI/Oアクセス性能を向上させてDBMSの性能を向上させる。

【0005】特開平9-274544号公報（文献2）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理的記憶装置上に配置する記憶装置において、前記論理的記憶装置の物理的記憶装置への配置を動的に変更することにより記憶装置のアクセス性能を向上する技術について開示している。アクセス頻度が高い物理記憶装置に記憶されているデータの一部を前記の配置動的変更機能を用いて他の物理記憶装置に移動することにより、特定の物理記憶装置のアクセス頻度が高くならないようにし、これにより記憶装置を全体としてみたときの性能を向上させる。また、配置動的変更機能による高性能化処理の自動実行方法についても開示している。

【0006】論文“高性能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価”（向井他、第11回データ工学ワークショップ（DEWS2000）論文集講演番号3B-3、2000年7月発行CD-ROM、主催：電子情報通信学会データ工学研究専

門委員会）（文献3）では、記憶装置の高機能化によるDBMSの性能向上について、リレーショナルデータベース管理システム（RDBMS）を用いたDBを例にして論じている。

【0007】記憶装置に対してアプリケーションレベルの知識としてRDBMSにおける問い合わせ処理の実行時の処理の実行プランを与えた場合、記憶装置は、RDBMSのある表に対する索引を読んだ後、その表のデータを記憶するどのブロックにアクセスすべきなのか判断できるようになる。そこで、索引を連続的にアクセスし、その索引によりアクセスすべき表のデータを保持するブロック群を把握し、それらに対するアクセスを効果的にスケジューリングすることによりデータの総アクセス時間を短縮することができる。

【0008】また、この処理はDBMSが実行されている計算機から独立に実行可能であり、計算機からの命令を待つ必要がない。また、データが複数台の物理記憶装置に分散配置されている場合にはそれぞれの物理記憶装置を並列にアクセス可能であり、よりDBMSの処理実行時間の短縮が期待できる。

【0009】文献3においては、前述の効果を擬似実験により確認している。擬似実験においては、実際に記憶装置上に前述の機能を実装するのではなく、ホスト側から先読みの指示を出す方法を採用している。実験に用いた記憶装置は2つのSCSIポートを保持し、記憶装置内のデータキャッシュは双ポートで共有している。そこで、片方でDBMSが実際に処理するデータをアクセスし、もう片方のポートで先読みするブロックの読み出しをすることにより、データの先読みを実現している。先読みするブロックはアクセストレース情報を基に決定している。

【0010】データの先読み（プリフェッチ）機能に関しては、以下のような技術が存在する。SCSI-2詳細解説（菅谷著、CQ出版社、ISBN4-7898-3523-5）の177ページ（文献4）には、SCSI-2規格として定義されているPre-Fetchコマンドについての説明が記述されている。このコマンドは、記憶装置に対して、指定された論理ブロックアドレスから指定されたデータ長分のデータを記憶媒体から読み出してデータキャッシュメモリに格納することを指示するコマンドである。

【0011】米国特許5887151（文献5）においては、1つのプリフェッチコマンドにおいて、プリフェッチすべき複数のブロックをリストとして受け取ることが可能な記憶装置に関する技術が公開されている。

【0012】論文“*Informed Prefetching and Caching*”（R. H. Patterson他著、Proc. of the 15th ACM Symp. on Operating System Principles, pp. 79-9

5、1995年12月）（文献6）においては、アプリケーションが発行する今後アクセスを行うファイルとアクセス先領域に関するヒントを用いて、計算機のOSにおいて計算機上のファイルキャッシュにデータをプリフェッチする機能とその制御方法について論じている。特に、既存のRDBMSを修正したプログラムを用いた評価も行われており、本技術のRDBMSに対する有効性も示されている。

【0013】

【発明が解決しようとする課題】従来の技術には以下のような問題が存在する。

【0014】文献1に記載の技術はDBMSにおける技術であり、文献の実施例中では記憶装置に対して相対ブロックアドレスを用いてアクセスしている。現在の大容量な記憶装置においては、記憶装置内にキャッシュメモリや複数の物理記憶装置を保持し、それらを制御する記憶装置制御装置により1つの記憶装置として動作しているものが存在する。このような記憶装置においては、内部制御の最適化によりより高いアクセス性能を発揮させる余地が存在するが、文献1の技術でDBMSはこのような記憶装置の内部処理の最適化までは考慮していない。

【0015】文献2に記載の技術においては、DBMSの動作特性を考慮していない。そのため、DBMSにより同時にアクセスされるデータを同じ物理記憶装置に配置してしまう可能性があり、これはDBMSのアクセス性能を低下させる要因となる。また、データの先読み等の記憶装置内キャッシュ制御の最適化については何も考慮していない。

【0016】文献3に記載の記憶装置の高機能化によるDBMSの性能向上について論じている部分においては、記憶装置がアプリケーションレベルの知識のとしてのRDBMSにおける問い合わせ処理の実行時の処理の実行プランを受け取る方法と、それを受け取った後にどのようなデータを用いてどのような処理を行う必要があるかの明確化が不十分である。

【0017】文献3、文献4、文献5に記載の記憶装置外部からプリフェッチすべきブロックを教える技術に関しては、プリフェッチコマンドによるアクセスの実行優先順位が明確でない問題点が存在する。同一の物理記憶装置に記憶されているブロックに対する異なるプリフェッチコマンドが記憶装置に到着した場合、どちらを優先的に実行すべきかは状況により異なるが、ブロック指定のプリフェッチコマンドではどのブロックを優先すべきか明確でない。

【0018】文献6に記載の技術に関しては、アプリケーションがヒントを発行する必要があるが、既存のDBMSに対して適用する際にはプログラムの修正を要する。一般に、DBMSに対しては極めて高い信頼性が要求される。DBMSは複雑なプログラムであり修正は容易で

ないことやプログラムの修正は信頼性低下の要因であることを考えると、この技術は未対応な既存DBMSに対して必ずしも適用できるものではない。

【0019】本発明の第一の目的は、DBMSが管理するデータを保持する記憶装置において、DBMS向けのアクセスの最適化を実施する記憶装置を実現することである。この記憶装置を用いることにより、既存のDBMSに対してプログラムの修正無しにDBMS稼動システムの性能を向上させることができるようになる。

【0020】本発明の第二の目的は、DBMSが管理するデータを保持する記憶装置において、DBのデータや処理に対する処理優先度を考慮したアクセスの最適化を実施する記憶装置を実現することである。DB毎や処理毎の処理優先度を考慮することにより、特定のDBに対する処理性能を保持するようなDBシステムを実現できるようにする。

【0021】

【課題を解決するための手段】一般に、RDBMSは、与えられた問い合わせ処理（クエリ）を実施する前にその実行プランを作成し、それに従って処理を実施する。このクエリ実行プランは、どのデータをどのようにアクセスし、アクセス後にホスト上でどのような処理を実行するかの手順を示すものである。そこで、DBMSの処理の実行情報である問い合わせ処理の実行プランを記憶装置自身が取得し、その情報と記憶装置内のデータ配置を考慮してこれからアクセスが行われるデータとそのアクセス順を予測し、それらのデータをあらかじめキャッシュメモリにプリフェッチしておくことにより、高いアクセス性能を有する記憶装置を実現する。

【0022】DBMSが直接、クエリ実行プランを記憶装置に対して与えることができない場合、同じ問い合わせ処理を実施する場合に作成されるクエリ実行プランは同じものとなることを利用する。多くのRDBMSには与えられた問い合わせに対する実行プランを外部に出力する機能を有しており、問い合わせを実行する前にその機能によりクエリ実行プランを取得し、その後にRDBMSに対して問い合わせを実行するようなプログラムを作成し、これを介して問い合わせ処理を実施するようにする。

【0023】また、DBのデータに対する処理優先度に関する情報を事前に取得しておき、プリフェッチ実行時にこれを考慮したプリフェッチやアクセス制御を実施する。特に、複数のDBに関するデータが同一の記憶装置上に存在する場合に、処理により高い優先度を要求するDBのデータに対して、より多くのキャッシュメモリとより高い物理記憶装置に対するアクセス割合を割り当て、高優先度DBに対する処理性能の劣化を防ぐ。

【0024】更に、クエリ実行プランを取得する際に、処理の優先度を指定するようにし、データに対する処理の優先度と同様に処理の優先度も考慮することにより高

い優先度を持つ処理の性能の劣化を防ぐ。

【0025】

【発明の実施の形態】以下、本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

＜第一の実施の形態＞本実施の形態では、DBMSが実行される計算機とデータキャッシュを保持する記憶装置が接続された計算機システムにおいて、記憶装置がDBMSに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、DBMSで実行されるクエリの実行プラン情報、DBの処理優先順位情報を取得し、それらを用いて記憶装置がより好ましいアクセス性能を提供する。

【0026】記憶装置は、DBMSに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、DBMSで実行されるクエリの実行プランを利用することにより、DBMSがこれからどのデータをどの順序でどのようにアクセスするかを把握することができる。そこで、この把握したアクセス方法に関する情報を利用して、あらかじめ利用される可能性が高いデータを記憶装置上のデータキャッシュ上に用意しておくことにより、DBMSに対してより高いアクセス性能を提供する。また、DBの処理優先順位情報を利用し、処理優先度が高いDBのデータや処理に対して、記憶装置が保持する物理記憶装置へアクセスを優先的に実施したり、また、データキャッシュの利用量をより多く割り当てたりすることにより、処理優先度が高いDBのデータや処理に対するアクセス性能を向上させる。

【0027】図1は、本発明の第一の実施の形態における計算機システムの構成図である。本実施の形態における計算機システムは、DBホスト80a、80b、DBクライアント81、処理性能管理サーバ82、記憶装置10から構成される。DBホスト80a、80b、DBクライアント81、処理性能管理サーバ82、記憶装置10はそれぞれが保有するネットワークインターフェイス78を通してネットワーク79に接続されている。また、DBホスト80a、80b、記憶装置10はそれぞれが保有するI/Oバスインターフェイス70からI/Oバス71を介してI/Oバススイッチ72に接続され、これらを通して記憶装置10とDBホスト80a、80b間のデータ転送を行う。

【0028】本実施の形態においては、記憶装置10とDBホスト80a、80b間のデータ転送を行うI/Oバス71とネットワーク79を異なるものとしているが、例えばiSCSIのような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されており、本実施の形態においてもこの技術を利用してよい。このとき、記憶装置10とDBホスト80a、80bにおいてI/Oバスインターフェイス70が省かれ、計算機システム内からI/Oバス71とI/Oバススイ

ッチ72が省かれる構成となる。

【0029】記憶装置10は、記憶領域を提供するもので、その記憶領域は記憶領域管理単位であるボリュームを用いて外部に提供し、ボリューム内の部分領域に対するアクセスや管理はブロックを単位として実行する。記憶装置10は、ネットワークインターフェイス78、I/Oバスインターフェイス70、記憶装置制御装置12、ディスクコントローラ16、物理記憶装置18から構成され、ネットワークインターフェイス78、I/Oバスインターフェイス70、記憶装置制御装置12、ディスクコントローラ16はそれぞれ内部バス20により接続され、ディスクコントローラ16と物理記憶装置18は物理記憶装置バス22により接続される。記憶装置制御装置12は、CPU24とメモリ26を有する。

【0030】メモリ26上には、記憶装置におけるキャッシュメモリとして利用するデータキャッシュ28が割り当てられ、記憶装置を制御するためのプログラムである記憶装置制御プログラム50が記憶される。また、メモリ26上には、物理記憶装置18の稼動情報である物理記憶装置稼動情報32、データキャッシュ28の管理情報であるデータキャッシュ管理情報34、DBMS110a、110bが利用するDBデータに対する処理優先度とそれらを考慮したディスクI/Oの管理情報である処理優先度付ディスクI/O管理情報36、DBホスト80a、80bで実行されているDBMS110a、110bで実行されるクエリの実行プランやそれを用いたプリフェッチの実行管理情報であるDBMS実行情報38、DBホスト80a、80bで実行されているDBMS110a、110bに関する情報であるDBMSデータ情報40、記憶装置10が提供するボリュームを物理的に記憶する物理記憶装置18上の記憶位置の管理情報であるボリューム物理記憶位置管理情報42を保持する。

【0031】図中の記憶装置10は、複数の物理記憶装置18を有し、1つのボリュームに属するデータを複数の物理記憶装置18に分散配置することが可能である。記憶装置制御プログラム50は、ディスクコントローラ16の制御を行うディスクコントローラ制御部52、データキャッシュ28の管理を行うキャッシュ制御部54、記憶装置10が提供するボリュームを物理的に記憶する物理記憶装置18上の記憶位置の管理に関する処理を行う物理記憶位置管理部56、I/Oバスインターフェイス70の制御を行うI/Oバスインターフェイス制御部58、ネットワークインターフェイス78の制御を行うネットワークインターフェイス制御部60を含む。

【0032】DBホスト80a、80b、DBクライアント81、処理性能管理サーバ82においては、それぞれCPU84、ネットワークインターフェイス78、メモリ88を有し、メモリ88上にオペレーティングシステム(OS)100が記憶・実行されている。

【0033】DBホスト80a、80bはI/Oバスインターフェイス70を有し、記憶装置10が提供するボリュームに対してアクセスを実行する。OS100内にファイルシステム104と1つ以上のボリュームからホストが利用する論理的なボリュームである論理ボリュームを作成するボリュームマネージャ102と、ファイルシステム104やボリュームマネージャ102により、OS100によりアプリケーションに対して提供されるファイルや論理ローボリュームに記憶されたデータの記録位置等を管理するマッピング情報106を有する。OS100が認識するボリュームやボリュームマネージャ102により提供される論理ボリュームに対して、アプリケーションがそれらのボリュームをファイルと等価なインターフェイスでアクセスするための機構であるローデバイス機構をOS100が有していても良い。

【0034】図中の構成ではボリュームマネージャ102が存在しているが、本実施の形態においてはボリュームマネージャ102における論理ボリュームの構成を変更することはないので、ボリュームマネージャ102が存在せずにファイルシステムが記憶装置10により提供されるボリュームを利用する構成に対しても本実施の形態を当てはめることができる。

【0035】DBホスト80a、80bのそれぞれのメモリ88上ではDBMS110a、110bが記憶・実行されている。DBMS110a、110bは内部にスキーマ情報114を有している。図中では、DBMS110a、110bが1台のホストに1つのみ動作しているが、後述するように、DBMS110a、110b毎の識別子を用いて管理を行うため、1台のホストにDBMSが複数動作していても本実施の形態に当てはめることができる。

【0036】DBホスト80a上ではDBMS情報取得・通信プログラム118とクエリプラン取得プログラム120が動作している。一方、DBホスト80b上ではDBMS情報取得・通信プログラム118とクエリプラン取得プログラム120が提供する機能をDBMS110b中のDBMS情報収集・通信部116が提供する。

【0037】DBクライアント81のメモリ88上では、DBMS110a、110bに対して処理要求を発行するDBMSフロントエンドプログラム126が記憶・実行される。図中では、DBMSフロントエンドプログラム126はDBホスト80a、80bと異なる計算機上で動作しているが、DBホスト80a、80b上で動作していても本実施の形態に当てはめることができる。

【0038】処理性能管理サーバ82のメモリ88上ではホスト情報設定プログラム130と処理性能管理プログラム132が記憶・実行される。図中では、ホスト情報設定プログラム130と処理性能管理プログラム132はDBホスト80a、80b、DBクライアント81

と異なる計算機上で動作しているが、それぞれ任意のDBホスト80a、80b、DBクライアント81上で動作していても本実施の形態に当てはめることができる。

【0039】図2はDBホスト80a、80bのOS100内に記憶されているマッピング情報106を示す。マッピング情報106中には、ボリュームローデバイス情報520、ファイル記憶位置管理情報530と論理ボリューム構成情報540が含まれる。ボリュームローデバイス情報520中にはOS100においてローデバイスを指定するための識別子であるローデバイスパス名521とそのローデバイスによりアクセスされる記憶装置10が提供するボリュームあるいは論理ボリュームの識別子であるローデバイスボリューム名522の組が含まれる。

【0040】ファイル記憶位置情報530中には、OS100においてファイルを指定するための識別子であるファイルパス名531とそのファイル中のデータ位置を指定するブロック番号であるファイルブロック番号532とそれに対応するデータが記憶されている記憶装置10が提供するボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名533とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号534の組が含まれる。

【0041】論理ボリューム構成情報540中にはボリュームマネージャ102により提供される論理ボリュームの識別子である論理ボリューム名541とその論理ボリューム上のデータの位置を示す論理ボリュームブロック番号542とその論理ブロックが記憶されているボリュームの識別子であるボリューム名501とボリューム上の記憶位置であるボリュームブロック番号512の組が含まれる。マッピング情報106を取得するには、OS100が提供している管理コマンドの実行や情報提供機構の利用、場合によっては参照可能な管理データの直接解析等を行う必要がある。

【0042】図3はDBMS110a、110b内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報114を示す。スキーマ情報114には、表のデータ構造や制約条件等の定義情報を保持する表定義情報551、索引のデータ構造や対象である表等の定義情報を保持する索引定義情報552、利用するログに関する情報であるログ情報553、利用する一時表領域に関する情報である一時表領域情報554、管理しているデータのデータ記憶位置の管理情報であるデータ記憶位置情報555とデータをアクセスする際の並列度に関する情報である最大アクセス並列度情報557を含む。

【0043】データ記憶位置情報555中には、表、索引、ログ、一時表等のデータ構造の識別子であるデータ構造名561とそのデータを記憶するファイルまたはローデバイスの識別子であるデータファイルパス名562と

その中の記憶位置であるファイルブロック番号563との組が含まれる。最大アクセス並列度情報557には、データ構造名561と、そのデータ構造にアクセスする際の一般的な場合の最大並列度に関する情報である最大アクセス並列度569の組が含まれる。スキーマ情報114を外部から取得するには、管理ビューとして外部に公開されているものをSQL等のデータ検索言語を用いて取得したり、または、専用の機構を用いて取得したりすることができる。

【0044】図4は記憶装置10内に保持されているボリューム物理記憶位置管理情報42を示す。ボリューム物理記憶位置管理情報42中には、ボリューム名501とそのボリューム上のデータ記憶位置であるボリューム論理ブロック番号512とその論理ブロックが記憶されている物理記憶装置18の識別子である物理記憶装置名502と物理記憶装置18上の記憶位置である物理ブロック番号514の組のデータが含まれる。

【0045】図5に記憶装置10内に保持されている物理記憶装置稼働情報32を示す。物理記憶装置稼働情報32中には、記憶装置10が提供するボリュームの識別子であるボリューム名501とそのボリューム名501を持つボリュームのデータを保持する物理記憶装置18の識別子である物理記憶装置名502、そしてボリューム名501を持つボリュームが物理記憶装置名502を持つ物理記憶装置18に記憶しているデータをアクセスするための稼働時間のある時刻からの累積値である累積稼働時間503、稼働率594計算のために前回利用した累積稼働時間503の値である旧累積稼働時間593とある一定時間内の動作時間の割合を示す稼働率594の組と、稼働率594計算のために前回累積稼働時間を取得した時刻である前回累積稼働時間取得時刻595を含む。

【0046】ディスクコントローラ制御部52はディスクコントローラ16を利用して物理記憶装置18へのデータアクセスする際の開始時刻と終了時刻を取得し、そのアクセスデータがどのボリュームに対するものかを判断して開始時刻と終了時刻の差分を稼働時間として対応するボリューム名501と物理記憶装置名502を持つデータの組の累積稼働時間503に加算する。ディスクコントローラ制御部52は一定間隔で以下の処理を行う。累積稼働時間503と旧累積稼働時間593、前回累積稼働時間取得時刻595と現データ取得時刻を用いて前回累積稼働時間取得時刻595と現データ取得時刻間の稼働率594を計算・記憶する。その後、取得した累積稼働時間503を旧累積稼働時間593に、現データ取得時刻を前回累積稼働時間取得時刻595に記憶する。

【0047】図6に記憶装置10内に保持されているDBMSデータ情報40を示す。DBMSデータ情報40中には、DBMSスキーマ情報711、データ構造物理

記憶位置情報712を含む。

【0048】DBMSデータ情報40中に含まれるデータは、DBホスト80a、80b上に存在するデータを利用する必要があるものが含まれる。記憶装置10は記憶装置10の外部に存在する情報を処理性能管理サーバ82で動作するホスト情報設定プログラム130を利用して取得する。ホスト情報設定プログラム130はネットワーク79を通し、DBホスト80a上で実行され、マッピング情報106等必要となる情報の収集処理を実施するDBMS情報取得・通信プログラム118や、DBホスト80b上で実行されているDBMS110b中のDBMS情報取得・通信プログラム118と等価な機能を実現するDBMS情報収集・通信部116を利用して必要な情報を収集する。

【0049】ホスト情報設定プログラム130は情報取得後、必要ならば記憶装置10に情報を設定するためのデータの加工を行い、ネットワーク79を通して記憶装置10に転送する。記憶装置10においては、ネットワークインターフェイス制御部60が必要な情報が送られてきたことを確認し、キャッシュ制御部54に渡し、必要な加工を行った後にその情報をDBMSデータ情報40中の適切な場所に記憶する。

【0050】前述のように、ホスト情報設定プログラム130は任意のDBホスト80a、80b上で実行されてもよい。あるいは、キャッシュ制御部54がホスト情報設定プログラム130の情報収集機能を有してもよい。これらの場合は、DBホスト80a、80bから情報を転送する際にI/Oパス71を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリュームを記憶装置10はDBホスト80a、80bに提供し、そのボリュームに対する書き込みがあった場合にI/Oパスインターフェイス制御部は情報の転送があったと判断し、その情報をキャッシュ制御部54に渡し、必要な加工を行った後にその情報をDBMSデータ情報40中の適切な場所に記憶する。

【0051】情報の収集処理に関しては、記憶装置10が必要になったときに外部にデータ転送要求を出す方法と、データの変更があるたびに外部から記憶装置10に変更されたデータを送る方法の2種類ともに利用することができる。ただし、DBMS110a、110bにおけるクエリ実行プランに関しては、実行する処理が明らかになった時点で受け取る必要があるため、記憶装置10は処理性能管理プログラム132もしくはクエリプラン取得プログラム120もしくはDBMS110bが与えるものを受動的に受け取る必要がある。

【0052】図7にDBMSデータ情報40中に含まれるDBMSスキーマ情報711を示す。DBMSスキーマ情報711は、DBMSデータ構造情報621、DBMSデータ記憶位置情報622、DBMSパーティション化表・索引情報623、DBMS索引定義情報62

4、DBMSホスト情報626、DBMSホストマッピング情報627を含む。DBMSデータ構造情報621はDBMS110a、110bで定義されているデータ構造に関する情報で、DBMS110a、110bの識別子であるDBMS名631、DBMS110a、110b内の表・索引・ログ・一時表領域等のデータ構造の識別子であるデータ構造名561、データ構造の種別を表すデータ構造種別640、データ記憶位置情報から求めることができるデータ構造が利用する総データ量を示すデータ構造データ量641、そのデータ構造をアクセスする際の最大並列度に関する情報である最大アクセス並列度569の組を保持する。このとき、データ構造によっては最大アクセス並列度569の値を持たない。

【0053】DBMSデータ記憶位置情報622はDBMS名631とそのDBMSにおけるデータ記憶位置管理情報555であるデータ記憶位置管理情報638の組を保持する。DBMSパーティション化表・索引情報623は、1つの表や索引のある属性値により幾つかのグループに分割したデータ構造を管理する情報で、パーティション化されたデータ構造が所属するDBMS110a、110bの識別子であるDBMS名631と分割化される前のデータ構造の識別子であるパーティション元データ構造名643と分割後のデータ構造の識別子であるデータ構造名561とその分割条件を保持するパーティション化方法644の組を保持する。今後、パーティション化されたデータ構造に関しては、特に断らない限り単純にデータ構造と呼ぶ場合にはパーティション化後のものを指すものとする。

【0054】DBMS索引定義情報624には、DBMS名631、索引の識別子である索引名635、その索引のデータ形式を示す索引タイプ636、その索引がどの表のどの属性に対するものかを示す対応表情報637の組を保持する。DBMSホスト情報626は、DBMS名631を持つDBMS110a、110bがどのホスト上で実行されているかを管理するもので、DBMS名631とDBMS実行ホストの識別子であるホスト名651の組を保持する。

【0055】DBMSホストマッピング情報627はDBホスト80a、80bのOS100内に記憶されているマッピング情報106収集したもので、ホスト名651とそのホストにおけるマッピング情報106を保持するマッピング情報648の組からなる。DBMSホスト情報626はシステム構成情報で管理者が設定するものである。DBMSスキーマ情報711中のその他のデータはDBMS110a、110bが管理しているスキーマ情報114の中から必要な情報を取得して作成する。

【0056】図8にDBMSデータ情報40中に含まれるデータ構造物理記憶位置情報712を示す。データ構造物理記憶位置情報712はDBMS110a、110bに含まれるデータ構造が記憶装置10内でどの物理記

憶装置18のどの領域に記憶されるかを管理するもので、データ構造を特定するDBMS名631とデータ構造名561、そのデータ構造内のブロックの通番であるデータ構造ブロック通番716、その外部からのアクセス領域を示すボリューム名501とボリュームブロック番号512、その物理記憶装置18上の記憶位置を示す物理記憶装置名502と物理ブロック番号514の組を保持する。この情報は、DBMSスキーマ情報711内のDBMSデータ記憶位置情報622とDBMSホストマッピング情報627とボリューム物理記憶位置メイン情報510を参照して、対応する部分を組み合わせることにより作成する。

【0057】DBMS110a、110b毎にシーケンシャルアクセスの方法が定まっている。DBMS名631とデータ構造名561により特定されるデータ構造毎に、シーケンシャルアクセス時のアクセス順を保持するようにソートしたデータをデータ構造物理記憶位置情報712は保持する。そして、データ構造毎のシーケンシャルアクセス順にしたがってデータ構造ブロック通番716を割り当てる。ここでは、対象とするDBMS110a、110bの種類を絞り、あらかじめデータ構造物理記憶位置情報712を作成するプログラムがDBMS110a、110bにおけるシーケンシャルアクセス方法を把握し、シーケンシャルアクセス順でソートされたデータを作成する。

【0058】本実施の形態のDBMS110a、110bにおけるシーケンシャルアクセス方法は以下の方法に従うものとする。あるデータ構造のデータをシーケンシャルアクセスする場合に、データ構造が記憶されているデータファイル名562とファイルブロック番号563を昇順にソートしその順序でアクセスを実行する。その他にシーケンシャルアクセス方法の決定方法としては、データファイルを管理する内部通番とファイルブロック番号563の組を昇順にソートした順番にアクセスする方法等が存在し、それらを利用したシーケンシャルアクセス方法の判断を実施してもよい。

【0059】図9に記憶装置10内に保持されているDBMS実行情報38を示す。DBMS実行情報38中には、現在有効な実行情報を管理する実行情報ID管理情報800と取得したクエリ実行プランをもとに作成するデータのアクセス方法を保持するDBMS処理計画情報805、シーケンシャルアクセスに対するプリフェッチを実行する際の管理情報であるシーケンシャルプリフェッチ情報810、B-T r e e索引を解釈することによるプリフェッチを実行する際の管理情報であるB-T r e e索引プリフェッチ情報820、プリフェッチを実行する際のアクセスパターンを把握する際に利用するデータ構造アクセス情報825を含む。

【0060】実行情報ID管理情報800中には実行情報の識別子である実行情報ID801とその実行情報で

示される処理を実施するDBMS 110a, 110bの識別子であるDBMS名631、その実行情報を作成するのに用いたクエリプランの識別子であるクエリプランID802、そのクエリの処理優先度であるクエリ優先度803、その実行情報をもとに実行するプリフェッチ用のキャッシュの管理情報のIDであるDBMSキャッシュ管理ID811の組を有する。各エントリ値が有効値の場合にはそのエントリは有効でそうでない場合にはそのエントリは無効である。

【0061】DBMS処理計画情報805中には、実行情報ID801と実行情報内の管理のための実行情報内部通番806とクエリ実行プランから識別されるデータ構造名561、クエリ実行プランから識別されるクエリ実行プラン中のデータ構造へのアクセス順序である実行順序807、そのデータ構造へのアクセス方法であるアクセスタイプ808の組を保持する。この情報に関しては、後述するクエリ実行プランに関する情報をもとに設定する。

【0062】シーケンシャルプリフェッチ情報810中には、DBMS処理計画情報のエントリを特定する実行情報ID801と実行情報内部通番806、プリフェッチ用キャッシュの管理領域の識別子であるDBMSキャッシュ管理ID811、そのプリフェッチ管理情報によりプリフェッチが実行されるデータ構造の領域を示すデータ領域範囲812、どこまでプリフェッチを実行したかを管理するプリフェッチポインタ813、そのプリフェッチ領域へホストからの実アクセスの状況を示すアクセス状況814の組を含む。データ領域範囲812が示す領域は、データ構造ブロック通番716を用いて領域範囲を示す。アクセス状況814は、ホストからのアクセスが実行されていない“未実行”、ホストからのアクセスが存在した“実行”のいずれかを保持する。

【0063】B-Tre e索引プリフェッチ情報820中には、DBMS処理計画情報のエントリを特定する実行情報ID801と実行情報内部通番806、プリフェッチ用キャッシュの管理領域の識別子であるDBMSキャッシュ管理ID811、索引により選択するデータの条件を保持する索引アクセス条件821の組を保持する。

【0064】データ構造アクセス情報825中には、データ構造を特定するDBMS名631とデータ構造名561、データ構造毎の最も最近に行われたある一定数のアクセス履歴情報であるデータアクセス情報826を含む。データアクセス情報826はアクセス先としてデータ構造ブロック通番716を用いる。また、アクセス先とアクセスサイズは組にしてFIFOアルゴリズムで管理する。

【0065】記憶装置10は、データキャッシュをある一定サイズの領域であるセグメントと呼ぶ管理単位を用いて管理する。図10に記憶装置10内に保持されてい

るデータキャッシュ管理情報34を示す。データキャッシュ管理情報34中には、データキャッシュ34のセグメントの状態を示すキャッシュセグメント情報720とキャッシュセグメントの再利用対象選定に利用するキャッシュセグメント利用管理情報740と、DBMS実行情報38を利用したプリフェッチ実行のために割り当てられたキャッシュセグメントを管理するDBMSデータキャッシュ管理情報830を含む。

【0066】キャッシュセグメント情報720中には、セグメントの識別子であるセグメントID721と、そのセグメントに記憶されているデータ領域を示すボリューム名501とボリューム論理ブロック番号512、そして、セグメントの状態を示すステータス情報722、後述するセグメントの管理に利用するリストの情報であるリスト情報723を含む。

【0067】ステータス情報722が示すセグメントの状態としては、物理記憶装置18上にセグメント内のデータと同じデータが記憶されている“ノーマル”、セグメント内のみ最新のデータが存在する“ダーティ”、セグメント内のデータに対する書き込み要求発行中を示す“ライト”、セグメント内に有効なデータが存在しない“インバリッド”が存在する。リスト情報723には、現在そのセグメントが属するリストの識別子と、そのリストのリンク情報が記憶される。図中では、リストは双方向リンクリストであるとしている。

【0068】キャッシュセグメント利用管理情報740中には、キャッシュセグメントの再利用対象選定に利用する2種類の管理リストであるメインLRUリスト、再利用LRUリストの管理情報として、メインLRUリスト情報741、再利用LRUリスト情報743が、現在のキャッシュセグメント情報720中のステータス情報722がダーティに設定されている数を示すダーティセグメント数カウンタ746とDBMS実行情報38を利用したプリフェッチを実行のために割り当てられたキャッシュセグメント数を示すDBMSプリフェッチ用割り当てセグメント数747が記憶される。メインLRUリスト情報741、再利用LRUリスト情報743は、それぞれリストの先頭であるMRUセグメントID、最後尾であるLRUセグメントID、そのリストに属するセグメント数を記憶する。

【0069】DBMSデータキャッシュ管理情報830中には、エントリの識別子であるDBMSキャッシュ管理ID811とこのエントリにより管理されるリストの先頭セグメント、最後尾セグメント、セグメント数を表す先頭セグメントID831、最後尾セグメントID832、領域セグメント数833の組を保持する。領域セグメント数833の値でそのエントリが利用中かどうかを判断し、その値が0以上の場合には利用中と判断し、負の値の場合には未利用と判断する。

【0070】図11に記憶装置10内に保持されている

処理優先度付ディスクI/O管理情報36を示す。処理優先度付ディスクI/O管理情報36はデータ構造の処理優先度に関する情報であるDBMSデータ構造処理優先度情報840とデータ構造に対して割り当てられた処理優先度をもとにした制御を行う際に利用する処理情報設定情報850、物理記憶装置18へのデータアクセスを実行する際にディスクコントローラ16に対するI/O要求の発行管理に利用するディスクI/O実行管理情報860を含む。

【0071】DBMSデータ構造処理優先度情報840中には、データ構造を特定するDBMS名631とデータ構造名561、そのデータ構造に与えられたデフォルトの処理優先度である処理優先度841、DBMS実行情報38で管理される実行情報中のクエリ優先度803を加味したデータ構造の処理優先度である実効処理優先度842の組と、データ構造に属さないその他の一般データに対するアクセスに与えられる処理優先度である一般データ処理優先度845を保持する。

【0072】処理情報設定情報850中には、処理優先度841とその処理優先度に割り当てられる物理記憶装置18へのI/O数の比率である割り当てI/O比率851、その処理優先度を持つ領域におけるクエリ毎にプリフェッチ用に割り当てるキャッシュセグメント数を示す領域プリフェッチ割り当て量852、その処理優先度を持つと判断されたクエリに対するプリフェッチ用に割り当てるキャッシュセグメント数の最大値を示すクエリプリフェッチ最大量853の組を保持する。

【0073】ディスクI/O実行管理情報860中は、物理記憶装置18の識別子である物理記憶装置502とそこへのディスクI/O発行を管理するために利用するI/O実行管理情報861の組を含む。I/O実行管理情報861中は、読出しアクセス要求を溜めるキューの配列である読出しキュー配列863と書込みアクセス要求を溜めるキューである書込みキュー864、処理優先度毎のI/O数割り当てを管理するために用いる処理優先度別処理残数カウンタ866を含む。

【0074】読出しキュー配列863へのアクセスは、処理優先度831とホストアクセス要求に答えるためのものかプリフェッチ実施のためのものを示す指示子867によりアクセスすべきアクセスキューを指定する。処理優先度別処理残数カウンタ866中には、処理優先度831と対応する処理優先度におけるアクセス可能残数を示す処理残りカウント値868の組を有する。

【0075】続いて、クエリの実行プラン利用した記憶装置10における制御方式を説明する。DBMS110a、110bは、処理を与えた場合にその結果を得るために内部的にどのような処理がどのような順番で実行されるかを示すクエリ実行プランを外部に提供する機能を有する。この機能により取得したクエリ実行プランを記憶装置10に与えることにより、記憶装置10はこれか

らどのような処理がDBMS110a、110bで行われるかを把握することができる。その情報を用いてデータをあらかじめデータキャッシュ28にプリフェッチしたり、あるいは、処理実行中にもうアクセスされないデータキャッシュ28上のデータを把握してそれらのデータを保持しているキャッシュセグメントを優先的に再利用することにより、より高いアクセス性能を得る。

【0076】クエリ実行プランの例として図12中にクエリ871とその処理を実現するためにDBMS110aが作成したクエリ実行プラン872を図示したものを示す。図に示されるように、クエリ実行プラン872はクエリ871の結果を得るために内部で実行する細分化された処理をノードとする木構造で表現することが可能である。図中でデータは末端から幹の方向に向かって流れていく。処理ノード875a、875b、875c、875d、875e、875f、875g、875hはクエリ861で実行される細分化された処理を表し、枝876は処理間のデータの流れ関係を示す。処理グループ877a、877b、877cはDBMS110aにおいて同時に実行される可能性がある処理の組を表し、1つの処理グループに属する処理が完了するまで他の処理グループに属する処理が実行されることはない。

【0077】処理グループ内での処理は、処理グループ内で行われる処理の内容とそれらの処理に利用されるデータの流れによりその実行順序は定まる。クエリ実行計画872では、まず処理グループ877bにより、表T3が全走査される。続いて処理グループ877cにより、表T4の全走査が行われその結果をもとに属性値Mの値が100未満の組が選択される。この表の全走査と属性値Mによるデータの選択は並行して実行される。続いて、処理グループ877aの処理が行われる。処理グループ877bと処理グループ877cの処理結果のハッシュ結合処理を実施する。その結果を用いて索引Ind1-1を参照して表T1内の対応データを検索するネストループ結合を実施し、その結果から表T1の属性Bの値の総和を求める。処理グループ877a内の処理は並行して実施される。

【0078】図13はクエリ実行プラン872が作成されたときに、記憶装置10に与えられるクエリ実行プランに関する情報であるクエリプラン情報880を示す。クエリプラン情報880中には、そのクエリ実行プランを持つ処理が行われるDBMS110a、110bの識別子であるDBMS名631とそのクエリ実行プランの識別子であるクエリプランID802、そのクエリ実行プランを持つ処理に与えられた処理の優先度であるクエリ優先度803、クエリの実行プランに関する詳細情報を保持するクエリ実行プラン情報881が含まれる。

【0079】クエリ実行プラン情報881には、クエリ実行プラン872を木構造で表現した場合の処理ノード875の識別子883とその親の処理ノード875の識

別子884、ハッシュ結合・ネステッドループ結合・ソートマージ結合・全表走査・表アクセス・索引アクセス・フィルタ・ソート・総和演算等のその処理ノード875で実施されるノード処理内容885、処理ノード875でデータ構造に対するアクセスを実施する処理の場合のアクセス先のデータ構造名561であるアクセスデータ構造886、処理ノード875が所属する処理グループ877の処理グループ間の処理の実行順序を示す処理順序887、そして、結合演算における結合処理条件や索引アクセスにおけるデータ検索条件、データの並列アクセス時のデータ分割方法等の処理ノード875で実行される処理の詳細情報であるノード処理詳細888を含む。なお、アクセスデータ構造886はノード処理内容885がデータに対するアクセスではないエントリには無効値を入れる。ノード処理詳細888は必ずしも含まれなくともよい。

【0080】クエリプラン情報880を記憶装置10が受け取るときに、DBクライアント81上のDBMSフロントエンドプログラム126がDBMS110aに処理要求を出した場合の処理手順は以下になる。まず、DBMSフロントエンドプログラム126はネットワーク79を通してクエリプラン取得プログラム120に対してDBMS110a上での処理の実行を依頼する。このとき、クエリ優先度803も指定する。また、クエリプラン取得プログラム120は処理を依頼されたときに、DBMSフロントエンドプログラム126がどのようなものかの識別情報を作成する。

【0081】クエリプラン取得プログラム120は、DBMSフロントエンドプログラム126が依頼した処理に対するクエリ実行プラン872をDBMS110aから取得し、指定されたクエリ優先度803と処理を実行するDBMS110aのDBMS名631、DBMSフロントエンドプログラム126の識別情報、そして取得したクエリ実行プラン872とクエリプラン取得プログラム120が付加するクエリ実行プラン872の識別子であるクエリプランID802をネットワーク79を介して処理性能管理サーバ82上の処理性能管理プログラム132に送る。クエリ実行プラン872その他情報を受け取った処理性能管理プログラム132は、取得したクエリ優先度803とDBMSフロントエンドプログラム126の識別情報、その他処理優先度の設定をもとに記憶装置10に与えるクエリ優先度803を決定し、その他取得した情報を用いてクエリプラン情報880を作成し、ネットワーク79を通して記憶装置10に送信する。

【0082】クエリプラン取得プログラム120は、クエリ実行プラン872その他情報を処理性能管理プログラム132に送った後に、DBMSフロントエンドプログラム126が依頼した処理をDBMS110aに対して送る。その結果をクエリプラン取得プログラム120

が取得し、DBMSフロントエンドプログラム126に結果を返す。

【0083】その後、DBMS名631とクエリプランID802で識別されるクエリ実行プラン872により実行された処理が完了したことをネットワーク79を介して処理性能管理プログラム132に伝える。クエリ実行プラン872に対応する処理の完了報告を受けた処理性能管理プログラム132はDBMS名631とクエリプランID802で識別されるクエリプラン情報880に対応する処理が完了したことをネットワーク79を通して記憶装置10に伝える。

【0084】上記の例では、クエリプラン取得プログラム120はDBMS110aが動作するDBホスト80a上で動作しているが、任意の計算機、つまり、任意のDBホスト80a、80b、DBクライアント81上で動作してもよい。また、DBMSフロントエンドプログラム126がクエリプラン取得プログラム120の機能を含んでいてもよい。

【0085】DBMS110bにおいては、クエリプラン取得プログラム120の役割をDBMS110bのDBMS情報通信部106がその役割を果たす。DBクライアント81上のフロントエンドプログラム126がDBMS110bに処理要求を出した場合の処理手順は以下になる。まず、DBMSフロントエンドプログラム126はネットワーク79を通してDBMS110bへ処理の実行を依頼する。このとき、クエリ優先度803も指定する。また、DBMS110bは処理を依頼されたときに、DBMSフロントエンドプログラム126の識別情報を作成する。

【0086】DBMS110bはDBMSフロントエンドプログラム126が依頼した処理に対するクエリ実行プラン872を作成し、指定されたクエリ優先度803とDBMS110bのDBMS名631、DBMSフロントエンドプログラム126の識別情報、そして作成したクエリ実行プラン872とDBMS110bが付加するクエリ実行プラン872の識別子であるクエリプランID802をDBMS情報通信部106を利用してネットワーク79を介して処理性能管理サーバ82上の処理性能管理プログラム132に送る。

【0087】クエリ実行プラン872その他情報を受け取った処理性能管理プログラム130は、取得したクエリ優先度803とDBMSフロントエンドプログラム126の識別情報、その他処理優先度の設定をもとに記憶装置10に与えるクエリ優先度803を決定し、その他取得した情報を用いてクエリプラン情報880を作成し、ネットワーク79を通して記憶装置10に送信する。

【0088】DBMS110bは処理性能管理プログラム132に対して情報を送った後に依頼された処理を実行し、その結果をDBMSフロントエンドプログラム1

26に返す。その後、DBMS名631とクエリプランID802で識別されるクエリ実行プラン872により実行された処理が完了したことをDBMS情報通信部106を利用してネットワーク79を介して処理性能管理プログラム132に伝える。クエリ実行プラン872に対応する処理の完了報告を受けた処理性能管理プログラム132はDBMS名631とクエリプランID802で識別されるクエリプラン情報880に対応する処理が完了したことをネットワーク79を通して記憶装置10に伝える。

【0089】前述のように、処理性能管理プログラム132は計算機システム内の任意の計算機、つまり、任意のDBホスト80a、80b、DBクライアント81上で動作してもよい。これまで述べてきた方法では、クエリプラン取得プログラム120またはDBMS110bは一旦処理性能管理プログラム132へ情報を送り、そこでクエリプラン情報880を作成し記憶装置10に与えている。

【0090】その代わりに、クエリプラン取得プログラム120またはDBMS110bが直接クエリプラン情報880を作成し、それを記憶装置10へ送ってもよい。このとき、DBホスト80a、80bから情報を転送する際にI/Oパス71を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリュームを記憶装置10はDBホスト80a、80bに提供し、そのボリュームに対する書き込みがあった場合にI/Oパスインターフェイス制御部は情報の転送があったと判断する。

【0091】上記の方法では、記憶装置10が受け取る情報はクエリプラン情報880中のクエリの実行プランに関する情報はクエリ実行プラン情報881であるが、処理性能管理プログラム132でDBMS実行情報38に設定するDBMS処理計画情報805、シーケンシャルプリフェッチ情報810、B-Tre e索引プリフェッチ情報820の内容を作成し送ってもよい。このとき、実行情報ID801、DBMSキャッシュ管理ID811、プリフェッチポインタ813、アクセス状況814に入れる値は任意のものでよい。また、これらの情報の作成方法は記憶装置10内で設定する場合のものと同一方法を用い、それらについては後述する。

【0092】記憶装置10がネットワーク79通してあるいはI/Oパス71を通してクエリプラン情報880を受け取った際の処理を説明する。図14にクエリプラン情報880を受け取った際の処理フローを示す。これは、ネットワークインターフェイス78がクエリプラン情報880を受け取った場合にはネットワークインターフェイス制御部60が、I/Oパスインターフェイス70がクエリプラン情報880を受け取った場合にはI/Oパスインターフェイス制御部58がキャッシュ制御部54にクエリプラン情報880を受け取ったことを伝え

ることによりキャッシュ制御部54が処理を開始する。ステップ2001で処理を開始する。

【0093】ステップ2002では取得したクエリプラン情報880をもとにプリフェッチ可能領域を把握し、DBMS実行情報38中に必要な情報を設定する。まず、実行情報ID管理情報800に空きエントリを探して実行情報を保存するための実行情報ID801を取得し、そこにクエリプラン情報880をもとにDBMS名631、クエリプランID802、クエリ優先度803を設定し、DBMSデータキャッシュ管理情報830中の空きエントリを捜し、そのDBMSキャッシュ管理ID811を設定する。

【0094】そして、DBMSデータキャッシュ管理情報830中の選択したDBMSキャッシュ管理IDを持つエントリ中の領域セグメント数を0に設定し、そのエントリを利用中に設定する。なお、これ以降のクエリプラン情報880を取得した際の処理においては、DBMS名631はクエリプラン情報880の値とし、また、ここで取得した実行情報ID801とDBMS名631は本処理中で常に参照可能な状態にしておく。

【0095】続いて、クエリ実行プラン情報881をもとにDBMS処理計画情報805を設定する。この情報には、クエリ実行プラン情報881中でアクセスデータ構造886に有効値が入っているエントリを選択し、それらを処理順序887によって降順にソートする。更に、処理順序887に同じ値が入っているものに関しては、索引をアクセスした後に表のデータを読む等、プランノード名883とプラン親ノード名884、ノード処理内容885によってデータ間の依存関係を調べ、先にアクセスされるものを先に持ってくるようにソートする。

【0096】これらのソート結果にしたがって昇順に実行情報内部通番806が割り当てられるようにクエリ実行プラン情報881のエントリの内容をDBMS処理計画情報805に設定していく。データ構造名561にアクセスデータ構造886を、実行順序807には処理順序887の値を設定する。

【0097】アクセスタイプ808は以下のように設定する。まず、DBMS名631とアクセスデータ構造886からクエリ実行プラン情報881のエントリに対応するデータ構造の内容を把握する。データ構造が木構造索引であるものには、それに対応するアクセスタイプ808には“木構造索引”を、データ構造が表の場合には、そのクエリ実行プラン情報881中のデータ依存関係を調べ、木構造索引をアクセスした結果をもとにアクセスする場合には“索引参照”を、その他の場合には“シーケンシャル”を設定する。

【0098】次に、シーケンシャルプリフェッチ情報810を設定する。シーケンシャルプリフェッチ情報810は、直前に設定されたDBMS処理計画情報805の

エントリのうち、アクセスタイプ 808 に“シーケンシャル”と設定されたエントリに対して設定される。まず、アクセスタイプ 808 に“シーケンシャル”と設定されたエントリを取り出す。DBMS 処理計画情報 805 中の確認中のエントリからデータ構造名 561 を取り出す。DBMS 名 631 とデータ構造名 561 を索引名 635 とみなして DBMS 索引定義情報 624 を参照してデータ構造がビットマップ索引かどうかを確認し、そうである場合にはシーケンシャルプリフェッチ情報 810 に何も設定しない。

【0099】DBMS 110a, 110b においては、最大アクセス並列度 569 によりシーケンシャル順にデータ構造を等分し、それらを並列にアクセスするものとする。DBMS 名 631 とデータ構造名 561 を用いて DBMS データ構造情報 621 を参照してそのデータ構造における最大アクセス並列度 569 を求める。シーケンシャルプリフェッチ情報 810 中に、現在設定中の実行情報 ID 801 と実行情報内部通番 806 を持つエントリを取得した最大アクセス並列度 569 の値の分だけ作成する。DBMS 名 631 とデータ構造名 561 を用いてデータ構造物理記憶位置情報 712 を参照し、データ構造を記憶する全領域のデータ構造ブロック通番 716 を取得し、それを取得した最大アクセス並列度 569 の値で等分し、現在作成中のシーケンシャルプリフェッチ情報 810 のエントリのデータ領域範囲 812 にそれぞれ設定する。

【0100】そして、それぞれのエントリにおいて、プリフェッチポインタ 813 にデータ領域範囲 812 に設定された値の先頭部分を設定し、アクセス状況 814 には“未実行”を入れる。更に、DBMS データキャッシュ管理情報 830 中の空きエントリを捜し、その DBMS キャッシュ管理 ID 811 を設定する。そして、DBMS データキャッシュ管理情報 830 中の選択した DBMS キャッシュ管理 ID 811 を持つエントリの領域セグメント数 833 を 0 に設定する。

【0101】最後に、B-T re e 索引プリフェッチ情報 820 を設定する。B-T re e 索引プリフェッチ情報 820 は、直前に設定された DBMS 処理計画情報 805 のエントリのうち、アクセスするデータ構造が木構造のものに対して作成する。まず、アクセスタイプ 808 に“木構造索引”と設定されたエントリを取り出す。そのエントリの実行情報内部通番 806 を取り出し、先に対応が決定されたクエリ実行プラン情報 881 中のエントリを特定し、ノード処理詳細 888 の内容を確認する。

【0102】その内容が木構造索引をアクセスする際の条件を含まないときには、B-T re e 索引プリフェッチ情報 820 には何も設定しない。ノード処理内容詳細 888 の内容が木構造索引をアクセスする際の条件を含むとき、その条件が前の処理の結果に依存する場合には

B-T re e 索引プリフェッチ情報 820 に何も設定しない。依存しない場合には、B-T re e 索引プリフェッチ情報 820 にエントリを作成し、現在確認中の DBMS 処理計画情報 805 のエントリの実行情報 ID 801 と実行情報内部通番 805 を値を作成したエントリに設定する。

【0103】そして、木構造索引をアクセスする際の条件を索引アクセス条件 821 に設定する。そして、DBMS データキャッシュ管理情報 830 中の空きエントリを捜し、その DBMS キャッシュ管理 ID 811 を設定する。そして、DBMS データキャッシュ管理情報 830 中の選択した DBMS キャッシュ管理 ID 811 を持つエントリの領域セグメント数 833 を 0 に設定する。

【0104】ステップ 2003 では、プリフェッチ処理に利用するキャッシュセグメント数の計算とそれに付随する処理を行う。まず現在設定中の実行情報 ID 801 により実行情報 ID 管理情報 800 を参照してクエリ優先度 803 を求める。続いて、現在設定中の実行情報 ID 801 により DBMS 処理計画情報 805 を参照しアクセスが実行されるデータ構造名 561 の組を求める。求めたデータ構造名 561 すべてに対して、DBMS データ構造処理優先度情報 840 を参照して処理優先度情報 841 を求める。

【0105】この値と前に取得したクエリ優先度 803 をもとにあらかじめ定められたルールを用いて各データ構造のこの処理における処理優先度を求める。そして、DBMS データ構造処理優先度情報 840 中の実効処理優先度情報 842 を参照して求めた各データ構造のこの処理における処理優先度と比較して、この処理における処理優先度の方が高い場合には実効処理優先度情報 842 をこの処理における処理優先度の値に更新する。

【0106】実行情報 ID 801 を持つシーケンシャルプリフェッチ情報 810 の全エントリを検索し、実行情報 ID 801 と選択された各エントリにおける実行情報内部通番 806 からアクセスが実行される DBMS 処理計画情報 805 を参照してデータ構造名 561 を求め、このデータ構造に対応する先に求めたこの処理におけるデータ構造の処理優先度により処理優先度設定情報 850 から領域プリフェッチ割当量 852 を求め、そのシーケンシャルプリフェッチ情報 810 のエントリに対するプリフェッチ割り当て量とする。

【0107】続いて、実行情報 ID 801 を持つ B-T re e 索引プリフェッチ情報 820 の全エントリを検索し、実行情報 ID 801 と選択された各エントリにおける実行情報内部通番 806 からアクセスが実行される DBMS 処理計画情報 805 を参照してデータ構造名 561 を求め、このデータ構造に対応する先に求めたこの処理におけるデータ構造の処理優先度により処理優先度設定情報 850 から領域プリフェッチ割り当て量 852 を求め、その B-T re e 索引プリフェッチ情報 820 の

エントリに対するプリフェッチ割り当て量とする。

【0108】実行情報ID801を持つシーケンシャルプリフェッチ情報810とB-Tree索引プリフェッチ情報820の全エントリに対するプリフェッチ割り当て量の総和を求める。先に求めたこの処理におけるデータ構造の処理優先度のうちの最も優先度が高い値を求め、これを用いて処理優先度設定情報850を参照して求めるクエリプリフェッチ最大量853とプリフェッチ割当量の総和を比較し、値の小さな方をこのクエリにおけるプリフェッチ用キャッシュ容量の希望値とする。

【0109】ステップ2004では、先に求めた希望値の分だけプリフェッチ用キャッシュを割り当て可能か確認する。DBMSプリフェッチ用割り当てセグメント数747と先に求めた希望値の和があらかじめ定めた閾値未満か確認し、閾値未満の場合には割り当て可能と判断してステップ2005に進み、閾値以上の場合には割り当て不可能としてステップ2006に進む。

【0110】ステップ2005では、先に求めた希望値分のキャッシュセグメントを再利用LRUリストのLRU (Least Recently Used: 最も昔に使われた) 側から選択してそこから外し、それにあわせてキャッシュセグメント情報720と再利用LRUリスト情報743を更新する。選択したキャッシュセグメントから新たなリストを作成し、それらを実行情報ID管理情報800を参照して実行情報ID801に対応するDBMSキャッシュ管理ID811を求め、先に作成したリストの先頭セグメントID831と最後尾セグメントID832と領域セグメント数833に対応するエントリに記憶し、キャッシュセグメント情報720を更新し、これをプリフェッチに利用するキャッシュセグメントのプールとする。次にステップ2008に進む。

【0111】ステップ2006では、確保可能なキャッシュセグメントを再利用LRUリストから選択してそこから外し、それにあわせて再利用LRUリスト情報743とキャッシュセグメント情報720を更新する。選択したキャッシュセグメントから新たなリストを作成し、それらを実行情報ID管理情報800を参照して実行情報ID801に対応するDBMSキャッシュ管理ID811を求め、それに対応するエントリに先に作成したリストの先頭セグメントID831と最後尾セグメントID832と領域セグメント数833を記憶し、キャッシュセグメント情報720を更新し、これをプリフェッチに利用するキャッシュセグメントのプールとする。

【0112】ステップ2007では、クエリプラン情報880中のクエリ優先度803を求め、これより優先度が低い実行情報が存在するか実行情報ID管理情報800を参照して確認し、存在する場合にはそれらが確保しているキャッシュセグメントを現在設定中の実行情報用にまわす。現在設定中の実行情報よりクエリ優先度が低いものをクエリ優先度が低い順に確認し、実行情報ID

管理情報800中の対応するDBMSキャッシュ管理ID811により識別されるDBMSデータキャッシュ管理情報830のエントリにより管理されるプリフェッチ用キャッシュセグメントのプールを確認し、そこにセグメントが存在する場合にはそれらを現在確保しているプリフェッチ用キャッシュのプールに回し、それにあわせてDBMSデータキャッシュ管理情報830を更新する。

【0113】これを先に求めた希望値分のキャッシュセグメントが現在設定中の実行情報に対応するプリフェッチ用キャッシュのプールに貯まるまで、あるいは、クエリプラン情報880中のクエリ優先度803より優先度が低い実行情報中のプリフェッチ用キャッシュのプールに存在するものすべて収集したらこのステップの処理を完了し、ステップ2008に進む。このとき、現在設定中の実行情報のクエリ優先度803より優先度が低い実行情報中のプリフェッチ用キャッシュのプールに存在するものを集めても先に求めた希望値分まで貯まらない場合、更に、よりクエリ優先度が低い実行情報により既にプリフェッチが実行されたデータを保持するキャッシュセグメントを現在設定中の実行情報に対応するプリフェッチ用キャッシュのプールに持ち込む処理を追加してもよい。

【0114】ステップ2008では、現在設定中の実行情報ID801を持つエントリのうち実行順序807の値が最も早いエントリより示されるデータ構造に対するプリフェッチ用キャッシュを割り当てる。現在設定中の実行情報ID801を持つエントリのうち実行順序807の値が最も早いエントリの実行情報内部通番806を求める。対応する実行情報内部通番806は複数存在する可能性がある。実行情報ID801により実行情報ID管理情報800を参照して対応するDBMSキャッシュ管理ID811を求めてプリフェッチ用プールの情報を取得する。実行情報ID801と実行情報内部通番806に対応するシーケンシャルプリフェッチ情報810中のエントリに対しては、プールから最初に行うプリフェッチ量だけキャッシュセグメントを割り当て、シーケンシャルプリフェッチ情報810からそのエントリにおけるDBMSキャッシュ管理ID811を求め、それらプールとエントリに対応するDBMSデータキャッシュ管理情報830のエントリを更新する。

【0115】実行情報ID801と実行情報内部通番806に対応するB-Tree索引プリフェッチ情報820のエントリに対しては、1セグメント分の領域を割り当て、B-Tree索引プリフェッチ情報820からそのエントリにおけるDBMSキャッシュ管理ID811を求め、プールとそのエントリに対応するDBMSデータキャッシュ管理情報830のエントリを更新する。このとき、すべてのプリフェッチ情報に対してプリフェッチ用のキャッシュセグメントが行き渡るようにし、プー

ル内のキャッシュセグメント量が当初に与えた値では足りなくなる場合には、シーケンシャルプリフェッチ情報810中の各エントリに割り当てるキャッシュ量を調節して全エントリでほぼ等しいキャッシュセグメント数が割り当てられるようにする。

【0116】ステップ2009では、プリフェッチ用のプールに空きがあるか確認する。まず、実行情報ID801により実行情報ID管理情報800を参照して対応するDBMSキャッシュ管理ID811を求めてプリフェッチ用プールの情報を取得する。そしてプリフェッチ用のプールにプリフェッチ用に未割り当てのキャッシュセグメントが存在するかどうかを確認し、存在する場合には、ステップ2010に進み、存在しない場合にはステップ2014に進む。

【0117】ステップ2010では、プリフェッチ可能なデータ構造のデータを保持する物理記憶装置18の稼働率がある閾値以上のものがあるか確認する。現在設定中の実行情報ID801を持つシーケンシャルプリフェッチ情報810とB-Tree索引プリフェッチ情報820のエントリでまだプリフェッチ用バッファが割り当てられていないものの実行情報内部通番806を求める。求めた実行情報ID801と実行情報内部通番806の組でDBMS処理計画情報805を参照し、プリフェッチ用キャッシュがまだ割り振られていないデータ構造のデータ構造名561を求める。

【0118】現在設定中の実行計画のDBMS名631と求めたデータ構造名561の組を用いてデータ構造物理記憶位置情報712を参照し、それらが記憶されているボリューム名501と物理記憶装置名502の組を用いる。求めたボリューム名501と物理記憶装置名502を用いて物理記憶装置稼働情報32を参照し、対応するエントリ中の稼働率594のうち最も最近のものを求め、その値があらかじめ定められた閾値以上の場合には検索したボリューム名501と物理記憶装置名502に対応するデータ構造に対してプリフェッチを早めに実施すべきであると判断する。確認をしたすべてのデータ構造に対して、そのようなものが1つでも見つかった場合にはステップ2011に進み、そうでない場合にはステップ2012に進む。

【0119】ステップ2011では、ステップ2010で早めにプリフェッチを実施すべきであると判断されたデータ構造に対してプリフェッチ用のキャッシュを割り当てる。ステップ2010において閾値を超える稼働率を保持するボリューム名501と物理記憶装置名502の組を求めるもとなったシーケンシャルプリフェッチ情報810またはB-Tree索引プリフェッチ情報820のエントリを求める。以下、ステップ2008と同じ方法で対応するエントリにプリフェッチ用キャッシュを割り当て、ステップ2012に進む。

【0120】ステップ2012では、ステップ2009

と同じ方法でプリフェッチ用のプールに空きがあるか確認をし、空きがある場合にはステップ2013に進み、ない場合にはステップ2014に進む。

【0121】ステップ2013においては、まだプリフェッチ用キャッシュが未割り当てのプリフェッチ情報にプリフェッチ用キャッシュを割り振る処理を行う。現在設定中の実行情報ID801を持つシーケンシャルプリフェッチ情報810とB-Tree索引プリフェッチ情報820のエントリでまだプリフェッチ用バッファが割り当てられていないものの実行情報内部通番806を求める。その中の最も小さな実行情報内部通番806に対応する実行順序807をDBMS処理計画情報805を参照して求める。

【0122】DBMS処理計画情報805で現在設定中の実行情報ID801と求めた実行順序807を持つエントリの実行情報内部通番806を求め、先に求めた現在設定中の実行情報ID801を持つシーケンシャルプリフェッチ情報810とB-Tree索引プリフェッチ情報820のエントリでまだプリフェッチ用バッファが割り当てられていないものの中でここで求めた実行情報内部通番806を持つエントリを選択する。それらのエントリにおいて、ステップ2008と同じ方法で対応するエントリにプリフェッチ用キャッシュを割り当てる。

【0123】以下、ステップ2009と同じ方法でプリフェッチ用のプールに空きがあるか確認をし、空きがある場合にはこの実行順序807をもとにしたプリフェッチ用キャッシュの割り当て処理をプリフェッチ用のプールに空きが無くなるまで繰り返し、ステップ2014に進む。

【0124】ステップ2014においては、これまでに割り振られたプリフェッチ用のキャッシュセグメントに対してプリフェッチアクセス先を決定する処理を行い、そのプリフェッチ先のデータがデータキャッシュ28上に既に存在するか確認し、それに応じた処理を実施する。現在設定中の実行情報ID801を持つシーケンシャルプリフェッチ情報810のエントリでプリフェッチ用キャッシュが割り当てられたものに対して、まずデータ領域範囲812の領域の先頭からキャッシュの割り当て分だけプリフェッチすることとする。そして、プリフェッチポインタ813を割り当て分だけ進める。エントリ中の実行情報ID801と実行情報内部通番806を求め、その組を用いてDBMS処理計画情報805を参照して対応するデータ構造名561を求める。求めたデータ構造名561と現在設定中のDBMS名631、先にデータ領域範囲812の先頭から割り当てたアクセス先をデータ構造ブロック通番716としてデータ構造物理記憶位置情報712を参照し対応するデータが記憶されているボリューム名501とボリューム論理ブロック番号512と物理記憶装置名502とその物理ブロック番号514を求める。

【0125】ボリューム名501とボリューム論理ブロック番号512を用いてキャッシュセグメント情報720を参照して、対応するブロックが既にデータキャッシュ28上に存在するかどうか確認する。データが存在する場合には、プリフェッチ要求を作成せず、その分のキャッシュセグメントはプールに戻す。そのデータを保持するセグメントが再利用LRUリスト中に存在する場合には、そのキャッシュセグメントとプリフェッチ用に割り当てられたセグメントとを交換する。これらのとき、キャッシュセグメント情報720と、必要に応じて再利用LRUリスト情報743とDBMSキャッシュ管理情報830中の対応エントリを更新する。

【0126】まだデータキャッシュ28上にデータが存在しない場合には、求めた物理記憶装置18と物理ブロック番号514と読み出し先のセグメントID721からプリフェッチアクセス要求を作成する。このアクセス要求にはアクセス先のDBMS名631とデータ構造名561を含めておく。

【0127】現在設定中の実行情報ID801を持つB-Tree索引プリフェッチ情報820のエントリでプリフェッチ用キャッシュが割り当てられたものに対しては、そのエントリが対応する木構造索引のルートを持つ部分を持つ部分をプリフェッチする。エントリ中の実行情報ID801と実行情報内部通番806を求め、その組を用いてDBMS処理計画情報805を参照して対応するデータ構造名561を求め、索引を特定する。

【0128】求めたデータ構造名561と現在設定中のDBMS名631からDBMSデータ記憶位置情報622とDBMSホスト情報626とDBMSホストマッピング情報627を参照し、木構造索引のルートデータを持つ領域のボリューム名501とボリューム論理ブロック番号512をDBMS110a、110bに関する知識を用いて求める。

【0129】求めたボリューム領域の情報を用いてキャッシュセグメント情報720を参照して、対応するブロックが既にデータキャッシュ28上に存在するかどうか確認する。データが存在する場合には、そのデータを解釈し、B-Tree索引プリフェッチ情報820のエントリ中の索引アクセス条件821とその他必要なマッピング情報を参照して次にアクセスするデータのボリューム名501とボリューム論理ブロック番号512を求める。

【0130】以下、データキャッシュ28上に存在しないデータに当たるまでデータの解釈による次アクセス先の把握とデータキャッシュ28上での存在確認を繰り返す。データ解釈可能範囲ですべてのデータがデータキャッシュ28上に存在する場合にはプリフェッチに割り当てたキャッシュセグメントをプリフェッチ用のプールに戻す処理を行なう。データキャッシュ28上に存在しないデータが見つかった場合には、そのデータのプリフェ

ッチを考える。前述したように、B-Tree索引プリフェッチ情報820のエントリにプリフェッチ用に割り当てられるキャッシュセグメント数は1である。木構造索引のルートデータがデータキャッシュ28上に存在した場合、データ解釈の結果としてアクセス先が複数の領域に分かれる可能性がある。

【0131】複数の領域をプリフェッチすることになった場合には、実行情報ID801により実行情報ID管理情報800を参照して対応するDBMSキャッシュ管理ID811を求めてプリフェッチ用プールの情報を取得し、プールに要求を満たすのに十分な空きキャッシュセグメントが存在するか確認する。十分に存在する場合は必要量を、そうでない場合は利用可能分すべてをB-Tree索引のプリフェッチ割り当て用に回し、DBMSデータキャッシュ管理情報830とキャッシュセグメント情報720の対応部分を更新する。すべてのプリフェッチ可能領域に対してプリフェッチ用のキャッシュを割り当てることができない場合には、データ解釈段数を距離として、ルートデータに近いものを優先してプリフェッチ用キャッシュを割り当てる。

【0132】プリフェッチ先が決定した後、それらのボリューム名501とボリューム論理ブロック番号512に対応する物理記憶装置名502とその物理ブロック番号514をボリューム物理記憶位置管理情報42を参照して求め、読み出し先のセグメントID721を指定してプリフェッチアクセス要求を作成する。このアクセス要求にはアクセス先のDBMS名631とデータ構造名561を含めておく。

【0133】このとき、シーケンシャルプリフェッチ情報810またはB-Tree索引プリフェッチ情報820をもとにプリフェッチ先を割り当てられたキャッシュセグメントに関して、キャッシュセグメント情報720中の対応するエントリのボリューム名501とボリューム論理ブロック番号512を割り当て先の領域を示すように更新し、ステータス情報722の値を“インバリッド”に設定する。

【0134】ステップ2015では、ステップ2014で作成されたプリフェッチアクセス要求をもとにプリフェッチアクセスを発行する。全てのプリフェッチアクセス要求に対して、その中のDBMS名631とデータ構造名561からDBMSデータ構造処理優先度情報840を参照して処理優先度841を求め、この値と現在設定中の実行情報のクエリ優先度803をもとにあらかじめ定められたルールを用いてこのデータ構造のこの処理における処理優先度を求める。

【0135】求めた処理優先度を処理優先度841とし指示子867に“プリフェッチ”を指定して、プリフェッチ先の物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863中の対応するキューに追加する。キューに追加されたアクセス要求

の物理記憶装置18へのアクセス発行制御については後述する。

【0136】ステップ2016でクエリプラン情報880を取得時の処理を完了する。

【0137】図15にクエリプラン情報880に対応するクエリの完了通知を受け取った時の処理フローを示す。これは、ネットワークインターフェイス78が受け取った場合にはネットワークインターフェイス制御部60が、I/Oバスインターフェイス70が受け取った場合にはI/Oバスインターフェイス制御部58がキャッシュ制御部54にクエリプラン情報880に対応するクエリの完了通知を受け取ったことを伝えることによりキャッシュ制御部54が処理を開始する。ステップ2601で処理を開始する。このとき、完了したクエリを識別するために、DBMS名631とクエリプランID802が与えられる。

【0138】ステップ2602では、与えられたDBMS名631とクエリプランID802を用いて実行情報ID管理情報800を参照して、対応する実行情報ID801を求める。

【0139】ステップ2603では、求めた実行情報ID801を用いてシーケンシャルプリフェッチ情報810とB-Tre e索引プリフェッチ情報820を参照し、対応するエントリを求める。対応するそれぞれのエントリ中のDBMSキャッシュ管理ID811を求め、それに対応するDBMSデータキャッシュ管理情報830中のエントリにより管理されるプリフェッチデータ管理リンクを取得し、その管理リンク中に存在するキャッシュセグメントを全て再利用LRUリストのMRU側に繋ぎ直し、それにあわせて再利用LRUリスト情報743とキャッシュセグメント情報720の対応部分を更新する。

【0140】そして、先に求めたDBMSキャッシュ管理ID811に対応するDBMSデータキャッシュ管理情報830のエントリの先頭セグメントID831と最後尾セグメントID832に無効値を、領域セグメント数833に-1を代入してそのエントリを無効化する。その後、先に求めた実行情報ID801に対応するシーケンシャルプリフェッチ情報810とB-Tre e索引プリフェッチ情報820中のエントリを全てクリアする。

【0141】ステップ2604では、先に求めた実行情報ID801に対応するDBMS処理計画情報805中のエントリを全てクリアする。

【0142】ステップ2605では、先に求めた実行情報ID801を用いて実行情報ID管理情報800中の対応するDBMSキャッシュ管理ID811を求め、それに対応するDBMSデータキャッシュ管理情報830中のエントリにより管理されるプリフェッチ用キャッシュのプールの管理リンクを取得し、その管理リンク中に

存在するキャッシュセグメントを全て再利用LRUリストのMRU側に繋ぎ直し、それにあわせて再利用LRUリスト情報743とキャッシュセグメント情報720の対応部分を更新する。そして、先に求めたDBMSキャッシュ管理ID811に対応するDBMSデータキャッシュ管理情報830のエントリをステップ2603と同様な方法で無効化する。その後、先に求めた実行情報ID801に対応する実行情報ID管理情報800のエントリをクリアする。

【0143】ステップ2606で処理を完了する。

【0144】図16に記憶装置10がI/Oバス71を通してDBホスト80a、80bから書き込みアクセス要求を受け取ったときの処理フローを示す。I/Oバス71を通してI/Oバスインターフェイス70に書き込み要求が到着した際、I/Oバスインターフェイス制御部58がキャッシュ制御部54に要求を伝え、キャッシュ制御部54は処理を開始する。ステップ2101で処理を開始する。このとき、DBホスト80a、80bからのアクセス先の指定は、ボリューム名501とボリューム論理ブロック番号512で示される。

【0145】ステップ2102では、データキャッシュ28上に旧データが存在するか確認する。これは、書き込み先のボリューム名501とボリューム論理ブロック番号512を用いてキャッシュセグメント情報720を参照し、対応エントリを求めることにより確認する。1セグメントでも存在する場合にはステップ2103に進み、存在しない場合にはステップ2106に進む。

【0146】ステップ2103では、存在した旧データを保持するキャッシュセグメントが、プリフェッチ管理リストに存在するかどうか確認する。これは、先に求めたキャッシュセグメント情報720のエントリ内のステータス情報722とリスト情報723を参照し、それが属するリストの識別子と状態を調べることにより確認する。1セグメントでも存在する場合にはステップ2104に進み、存在しない場合にはステップ2105に進む。

【0147】ステップ2104では、書き込み処理に伴う、プリフェッチ用キャッシュプールの調整を行う。書き込み先として、プリフェッチ管理リストに存在するセグメント数と同数のキャッシュセグメントを再利用LRUリストから選択して確保する。そして、書き込み先の旧データを保持するセグメント毎に、そのセグメントが属するリストの識別子からそれを管理するDBMSデータキャッシュ管理情報830中の管理領域を示すDBMSキャッシュ管理ID811を求め、それによりプリフェッチ用セグメントを管理するシーケンシャルプリフェッチ情報810またはB-Tre e索引プリフェッチ情報820の対応エントリを求め、そのエントリ中の実行情報ID801を用いて実行情報ID管理情報800を参照し、プリフェッチ用プールのDBMSキャッシュ管

理ID811を求める。

【0148】求めたプールの管理情報に先に確保したキャッシュセグメントを1つ追加する。これを書き込み先の旧データを保持するセグメントすべてについて実行し、対応するキャッシュセグメント情報720とDBMSデータキャッシュ管理情報830のエントリの値を更新し、ステップ2105に進む。

【0149】ステップ2105では、書き込み先の旧データを保持するセグメントを現在繋がれている管理リストから外し、書き込み先キャッシュとして割り当てる。これは、ステップ2103で求めたセグメントが属する管理リストの識別子からどのリストに属しているかを判断し、それに応じて対応するキャッシュセグメント情報720とメインLRUリスト情報741、再利用LRUリスト情報743、DBMSデータキャッシュ管理情報830の対応エントリの値を更新し、ステップ2106に進む。

【0150】ステップ2106では、まだ書き込み先キャッシュセグメントが割り当てられていないものにそれを割り当てる。まだ割り当てられていないデータの保持に必要な数のキャッシュセグメントを再利用LRUリストのLRU側から確保し、再利用LRUリスト情報743と対応するのキャッシュセグメント情報720のエントリをそれにあわせて更新する。

【0151】ステップ2107では、取得したデータキャッシュにDBホスト80a、80bから送られてきたデータを転送する。

【0152】ステップ2108では、書き込み後のキャッシュ管理情報の更新処理を行う。まず、旧データを保持していたキャッシュセグメントに関して、キャッシュセグメント情報720の対応エントリ中のステータス情報722からその状態を確認し、“ダーティ”であったものの数を求める。そして、書き込み先として利用したキャッシュセグメント数から先に求めた旧データに関するダーティ数を引いた分だけダーティセグメント数カウンタ672の値を増加させる。

【0153】そして、キャッシュセグメント情報720中の書き込み先セグメントに対応するエントリのステータス情報722を“ダーティ”に設定する。その後書き込み先として利用したキャッシュセグメントをメインLRUリストのMRU側に接続し、メインLRUリスト情報741と対応するのキャッシュセグメント情報720のエントリをそれにあわせて更新する。

【0154】ステップ2109ではI/Oバスインターフェイス制御部58に対してDBホスト80a、80bに書き込み処理が完了したことを報告することを依頼し、I/Oバスインターフェイス制御部58はI/Oバスインターフェイス70を利用してI/Oバス71を介してDBホスト80a、80bに処理完了を報告する。

【0155】ステップ2110で処理を完了する。

【0156】図17に記憶装置10がI/Oバス71を通してDBホスト80a、80bから読み出しアクセス要求を受け取ったときの処理フローを示す。I/Oバス71を通してI/Oバスインターフェイス70に読み出し要求が到着した際、I/Oバスインターフェイス制御部58がキャッシュ制御部54に要求を伝え、キャッシュ制御部54は処理を開始する。ステップ2201で処理を開始する。このとき、DBホスト80a、80bからのアクセス先は、ボリューム名501とボリューム論理ブロック番号512で示される。

【0157】ステップ2202では、データキャッシュ28上にデータが存在するか確認する。これは、読み出し先のボリューム名501とボリューム論理ブロック番号512を用いてキャッシュセグメント情報720を参照し、対応エントリが存在し、かつ、ステータス情報722がインバリッドでないか調べることにより確認する。全要求データに関してデータキャッシュ28上に存在する場合にはステップ2207に進み、そうでない場合にはステップ2203に進む。

【0158】ステップ2203では、アクセス先のデータを読み出すプリフェッチ処理が発行済みかどうか確認し、存在する場合にはそれをホストからの要求によるものに変更する。まず、アクセス先のボリューム名501とボリューム論理ブロック番号512を用いてボリューム物理記憶位置管理情報42を参照し、アクセス先領域に対応する物理記憶装置名502と物理ブロック番号514を求める。

【0159】これをもとに、ディスクI/O実行管理情報860を参照し、その中の物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863を参照し、その中で指示子867が“プリフェッチ”のキューに存在するアクセス要求の中に先に求めた読み出し要求に対応するものが存在するかどうか確認する。存在した場合には、それを同じ処理優先度831における指示子867が“ホスト要求読出”のキューに繋ぎなおす。

【0160】ステップ2204では、まだ読み出し先キャッシュセグメントが割り当てられていないものにそれを割り当てる。まだ割り当てられていないデータの保持に必要な数のキャッシュセグメントを再利用LRUリスト743のLRU側から確保する。再利用LRUリスト情報743と対応するのキャッシュセグメント情報720のエントリをそれにあわせて更新する。キャッシュセグメント情報720中の対応するエントリのボリューム名501とボリューム論理ブロック番号512を読み出し先を示すように更新し、ステータス情報722の値を“インバリッド”に設定する。

【0161】ステップ2205では、ステップ2204でキャッシュセグメントが割り当てられたものに対して読み出し要求を発行する。このとき、各キャッシュセグ

メント毎に、アクセス先のボリューム名501とボリューム論理ブロック番号512を用いてボリューム物理記憶位置管理情報42を参照し、それに対応する物理記憶装置名502と物理ブロック番号514を求める。

【0162】更に、アクセス先のボリューム名501とボリューム論理ブロック番号512を用いてデータ構造物理記憶位置情報712を参照して、その領域に対応するDBMS110a, 110bのデータ構造を示すDBMS名631とデータ構造名561を求める。DBMS名631とデータ構造名561が求まった場合には、その組を用いてDBMSデータ構造処理優先情報840を参照し、実効処理優先度842を求め、このアクセスの処理優先度とする。また、求まらなかった場合には、一般データ処理優先度845を参照し、その値をこのアクセスの処理優先度とする。

【0163】求めた情報をもとに読み出しのアクセス要求を作成する。その中には、物理記憶装置名502と物理ブロック番号514とデータを読み出し先のキャッシュのセグメントID721を含める。データを読み出すために求めたアクセスの処理優先度を処理優先度841とし指示子867に“ホスト要求読み出し”を指定して、アクセス先の物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置18へのアクセス発行制御については後述する。

【0164】ステップ2206では、発行した読み出しアクセス要求が全て完了し、DBホスト80a, 80bから読み出し要求があったデータが全て揃うまで待ち、全てのデータが揃ったらステップ2207に進む。

【0165】ステップ2207ではI/Oバスインターフェイス制御部58に対してDBホスト80a, 80bからの読み出し要求があったデータがキャッシュ上のどこに存在するのかの情報としてセグメントID721の組を返す。I/Oバスインターフェイス制御部58はI/Oバスインターフェイス70を利用してI/Oバス71を介してDBホスト80a, 80bにそのデータを転送し、DBホスト80a, 80bとの間の処理を完了する。

【0166】ステップ2208では、アクセス先のデータの種類を確認する。アクセス先のボリューム名501とボリューム論理ブロック番号512を用いてデータ構造物理記憶位置情報712を参照して、その領域に対応するDBMS110a, 110bのデータ構造を示すDBMS名631とデータ構造名561を求める。それが求まった場合にはDBデータであるとしてステップ2209に進み、求まなかった場合にはステップ2210に進む。

【0167】ステップ2209では、DBデータの読み出しアクセス後の処理を実施する。その処理の詳細は後

述する。処理が完了後、ステップ2211に進み通してDBホスト80a, 80bから読み出しアクセス要求を受け取ったときの処理を完了する。

【0168】ステップ2210では、アクセス先キャッシュセグメントをメインLRUリストのMRU側に接続し、対応するキャッシュセグメント情報720のエントリとメインLRUリスト情報741と必要であれば再利用LRUリスト情報743をそれにあわせて更新する。ステップ2211に進みDBホスト80a, 80bから読み出しアクセス要求を受け取ったときの処理を完了する。

【0169】図18と図19はDBデータの読み出しアクセス後の処理の処理フローを示す。この処理を開始する際には、アクセス先のボリューム名501とボリューム論理ブロック番号512とそのデータを保持するセグメントID721が与えられる。ステップ2301で処理を開始する。

【0170】ステップ2302では、まず、アクセス先のボリューム名501とボリューム論理ブロック番号512を用いてデータ構造物理記憶位置情報712を参照して、その領域に対応するDBMS110a, 110bのデータ構造を示すDBMS名631とデータ構造名561とその領域のデータ構造ブロック通番716を求める。求めたDBMS名631とデータ構造名561によりデータ構造アクセス情報825を参照し、データアクセス情報826に求めたデータ構造ブロック通番716から識別されるアクセス先を追加する。

【0171】ステップ2303では、DBMS索引定義情報624をDBMS名631とデータ構造名561を索引名635として参照し、アクセス先が木構造索引かどうか確認する。参照結果として、エントリが見つからなかった場合、あるいは索引タイプ636が木構造索引でない場合にはステップ2304に進み、索引タイプ636が木構造索引である場合にはステップ2311に進む。

【0172】ステップ2304では、データアクセス情報826からDBMS名631とデータ構造名561に対応するデータをコピーし、そのデータをアクセス先のデータ構造ブロック通番716を用いてソートする。

【0173】ステップ2305では、アクセス先がシーケンシャルアクセスによるものかどうか確認する。その判断には、ステップ2304のソート結果を用いる。ソート結果を用いてアクセス先の前の領域がほぼ連続してアクセスされているかを確認し、アクセスされていたらシーケンシャルアクセスの一部とし、そうでない場合には、シーケンシャルアクセスの一部ではないとする。

【0174】DBMS110a, 110bでビットマップ索引を用いた場合には、シーケンシャルアクセスに近いが、完全に連続していないアクセスが実行されることがある。そこで、シーケンシャル性の判断時には、必ず

しも完全に連続していなくともよいとする。つまり、アクセス先領域がある一定値以下の間隔で飛び飛びに存在する場合もシーケンシャルアクセスであると判断する。シーケンシャルアクセスの一部と判断される場合には、ステップ2306に進み、そうでない場合には、ステップ2339に進む。

【0175】ステップ2306では、アクセス先に対応する実行情報が存在するか確認する。先に求めたDBMS名631からそれに対応する実行情報を管理する実行情報ID801を求め、これと先に求めたデータ構造名561によりDBMS処理計画情報805を参照し、対応エントリを求める。そのエントリにおけるアクセスタイプ808を調べ、“シーケンシャル”であるものの実行情報ID801と実行情報内部通番806を求める。このとき、1つも対応するエントリが見つからない場合にはステップ2339に進み、そうでない場合にはステップ2307に進む。

【0176】ステップ2307では、シーケンシャルプリフェッチ情報810内に対応エントリが存在するか確認する。ステップ2306で求めた実行情報ID801と実行情報内部通番806の組を用いてシーケンシャルプリフェッチ情報810を参照し、対応するエントリの中で先に求めたアクセス先のデータ構造ブロック通番716がデータ領域範囲に入るものを求め、そのエントリ中のDBMSキャッシュ管理ID811を求め、これを用いてDBMSデータキャッシュ管理情報830を参照してこれにより管理されるセグメントのリストを求めて、そのリストに存在するキャッシュセグメントが保持しているデータのボリューム名501とブロック番号512を求める。

【0177】このプリフェッチされたデータ領域内にアクセス先のもが含まれるか確認する。このとき、アクセス先と求めたプリフェッチ済み領域が完全に一致していなくとも、その距離が一定ブロック数以下であれば含まれると判断する。この距離は、データ構造ブロック通番716の差で判断し、プリフェッチ済み領域の対応値はボリューム名501とブロック番号512を用いてデータ構造物理記憶位置情報712を参照することにより求めることができる。この判断により、含まれると判断された場合には、そのシーケンシャルプリフェッチ情報810のエントリがアクセス先に対応するエントリと判断してステップ2330に進み、そのようなものが見つからなかった場合には、ステップ2308に進む。

【0178】ステップ2308では、アクセス先データを保持するキャッシュセグメントを再利用LRUリストのMRU側に接続し、再利用LRUリスト情報743と対応するのキャッシュセグメント情報720のエントリそして必要に応じてメインLRU情報741をそれにあわせて更新する。

【0179】ステップ2309では、シーケンシャルプ

リフェッチ情報にアクセス先におけるシーケンシャルアクセスの情報を設定する。先に求めた、アクセス先に対応する実行情報ID801と実行情報内部通番806の組をもとに新規にシーケンシャルプリフェッチ情報810を設定する。なお、アクセス先に対応する実行情報ID801と実行情報内部通番806の組が複数存在する場合には、どれか1つを適当に選択する。

【0180】アクセス先が他のシーケンシャルプリフェッチ情報810内の既に設定されている同じ実行情報ID801と実行情報内部通番806を持つエントリのデータ領域範囲812に含まれる場合には、既存のエントリのデータ領域範囲812を現在確認中のアクセス先より前の部分のみとし、後半部分を今回新たに設定するシーケンシャルプリフェッチ情報810のエントリのデータ領域範囲812に割り当てる。そうでない場合には、データ領域範囲812は現在のアクセス先にあらかじめ定められた動的プリフェッチ領域拡張用のブロック数を足したものに設定する。このとき、他の同じ実行情報ID801と実行情報内部通番806を持つ既存エントリと重なる部分がある場合は現在設定中のデータ領域範囲812において重なり部分を切り捨てる。

【0181】プリフェッチポイント813を現在のアクセス先の次のブロック番号に設定し、アクセス状況814には“実行”を入れる。更に、DBMSデータキャッシュ管理情報830中の空きエントリを捜し、そのDBMSキャッシュ管理ID811を設定する。そして、DBMSデータキャッシュ管理情報830中の選択したDBMSキャッシュ管理IDを持つエントリの領域セグメント数833を0に設定し、ステップ2332に進む。

【0182】ステップ2330では、シーケンシャルプリフェッチ情報810のアクセス先に対応するエントリにおいて、シーケンシャルアクセス時に現在確認中のアクセス先より前にアクセスされるべき領域のデータをプリフェッチしていたものと確認中のアクセス先のもを実行情報のプリフェッチ用キャッシュのプールに戻す処理を行う。ステップ2307で既にそのエントリに対応するDBMSキャッシュ管理ID811とそれにより管理されるセグメントのリストは求まっているので、それからプールに戻すキャッシュセグメントを求める。

【0183】また、シーケンシャルプリフェッチ情報810の対応するエントリ中の実行情報ID801を用いて実行情報ID管理情報800を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するDBMSキャッシュ管理ID811を求め、先に求めたプールに戻すキャッシュセグメントをプールに追加する。そして、それにあわせて対応するDBMSデータキャッシュ管理情報830とキャッシュセグメント情報720のエントリを更新する。

【0184】ステップ2331では、対応するシーケンシャルプリフェッチ情報810のエントリを更新する。

アクセス先がプリフェッチプリフェッチポインタ813を追いついたらその値をアクセス先に1を加算したものに更新する。また、データ領域範囲812の拡張を、アクセス先がシーケンシャルプリフェッチ情報810のエントリ内のデータ領域範囲812の終端から一定の距離の場所まで進み、かつ、それに連続したデータ領域範囲812と同じ実行情報ID801と実行情報内部通番806を持つエントリがシーケンシャルプリフェッチ情報810に存在しない場合に実行する。この場合、あらかじめ定められた動的プリフェッチ領域拡張用のブロック数を最大値として他のエントリと重ならない分量をデータ領域範囲812の終端に足し、ステップ2332に進む。

【0185】ステップ2332においては、現在確認中の実行情報のプリフェッチ用キャッシュのプールの残量を確認する。シーケンシャルプリフェッチ情報810の対応するエントリ中の実行情報ID801を用いて実行情報ID管理情報800を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するDBMSキャッシュ管理ID811を求め、DBMSデータキャッシュ管理情報830を参照してそれに対応する領域セグメント数833を求める。その値があらかじめ定められた閾値以上の場合にはステップ2333に進み、そうでない場合にはステップ2342に進み処理を完了する。

【0186】ステップ2333では、現在のアクセス先に対応するシーケンシャルプリフェッチ情報810により管理される領域にプリフェッチすべきデータが存在するかどうか確認する。これは、エントリのプリフェッチポインタ813がデータ領域範囲812内に存在するかどうかを確認することにより判断でき、存在する場合には、プリフェッチすべきデータがあるとしてステップ2337に進み、存在しない場合にはプリフェッチすべきデータが存在しないとしてステップ2334に進む。

【0187】ステップ2334では、次にプリフェッチを実施すべきシーケンシャルプリフェッチ情報810内のエントリの検索と決定を行う。現在確認中のエントリに対応するシーケンシャルプリフェッチ情報810のエントリが保持する実行情報ID801によりシーケンシャルプリフェッチ情報810のエントリを検索し、そのDBMSキャッシュ管理情報ID811を求め、それを用いてDBMSデータキャッシュ管理情報830を参照してプリフェッチされたセグメント数が0、かつ、アクセス状況814が未実行なものを選択する。

【0188】選択されたシーケンシャルプリフェッチ情報810のエントリ中で実行情報内部通番806が最も小さいものを次にプリフェッチを行うエントリとして選択する。このとき、複数エントリが同じ実行情報内部通番806を持つ可能性があるが、その場合には、対象の中でデータ領域範囲812が示す値が最も小さなものを

選択する。

【0189】ステップ2335では、ステップ2334での次のプリフェッチ対象となるシーケンシャルプリフェッチ情報810のエントリの選択が成功したかを判断する。選択に成功した場合にはステップ2336に進み、既にプリフェッチすべきものが存在せず選択に失敗した場合にはステップ2342に進み処理を完了する。

【0190】ステップ2336では、プリフェッチ用の領域を割り当て、プリフェッチコマンドを発行する。まず、シーケンシャルプリフェッチ情報810の対応するエントリ中の実行情報ID801を用いて実行情報ID管理情報800を参照し、対応する実行情報のDBMSキャッシュ管理ID811を求め、DBMSデータキャッシュ管理情報830を参照して、プリフェッチ用キャッシュのプールを管理するリストを取得し、そこからプリフェッチ用に割り当てるキャッシュセグメントを確保する。プリフェッチ実行量はあらかじめ定められた量であるとする。

【0191】この確保したキャッシュセグメントをシーケンシャルプリフェッチ情報810の対応するDBMSデータキャッシュ管理ID811で示されるDBMSデータキャッシュ管理情報830で管理されるプリフェッチデータ管理リストに繋ぎ、DBMSデータキャッシュ管理情報830のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報720中の対応する部分を更新する。

【0192】確保したキャッシュセグメントにプリフェッチ先を割り当て、キャッシュセグメント情報720中の対応するエントリのボリューム名501とボリューム論理ブロック番号512を割り当て先の領域を示すように更新し、ステータス情報722の値を“インバリッド”に設定する。アクセス先は、プリフェッチ対象となったシーケンシャルプリフェッチ情報810の対応するエントリのプリフェッチポインタ813から連続して割り当てられたキャッシュ量分の領域である。プリフェッチポインタ813はデータ構造ブロック通番716に対応している。そして、プリフェッチポインタ813を割り当て分だけ進める。シーケンシャルプリフェッチ情報810の対応するエントリの実行情報ID801から実行情報ID管理情報800を参照して対応するDBMS名631を求め、エントリの実行情報ID801と実行情報内部通番806を用いてDBMS処理計画情報805を参照してデータ構造名561を求め、これらの値を用いてデータ構造物理記憶位置情報712を参照してアクセス先のアクセスする物理記憶装置18上の領域である物理記憶装置名502と物理ブロック番号514を求める。また、DBMS名631とデータ構造名561を用いてDBMSデータ構造処理優先情報840を参照し、実効処理優先度842を求め、このアクセスの処理優先度とする。

【0193】割り当てたキャッシュセグメントを用いてプリフェッチアクセス要求を作成する。その中には、物理記憶装置名502と物理ブロック番号514とデータを読み出し先のキャッシュのセグメントID721を含める。求めたアクセスの処理優先度を処理優先度841とし指示子867に“プリフェッチ”を指定して、アクセス先の物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置18へのアクセス発行制御については後述する。その後、ステップ2342に進み処理を完了する。

【0194】ステップ2337では、アクセス先に対応するシーケンシャルプリフェッチ情報810のエントリにおける既プリフェッチデータ量を確認する。この値は、シーケンシャルプリフェッチ情報810の対応するDBMSデータキャッシュ管理ID811によりDBMSデータキャッシュ管理情報830を参照し、対応する領域セグメント数833の値を求める。この値が既プリフェッチデータ量に対応し、この値がある閾値未満の場合にはステップ2338に進み、閾値以上である場合には、ステップ2342に進み処理を完了する。

【0195】ステップ2338では、アクセス先に対応するシーケンシャルプリフェッチ情報810のエントリで管理される領域をプリフェッチの実行対象に設定し、ステップ2336に進む。

【0196】ステップ2311では、現アクセス先データの中身を解釈することによりプリフェッチが可能かどうか確認する。先に求めたDBMS名631からそれに対応する実行情報を管理する実行情報ID801を求め、これと先に求めたデータ構造561によりDBMS処理計画情報805を参照し、対応エントリを求める。ここで対応エントリが求まらない場合には、プリフェッチ不可能としてステップ2339に進む。求まった場合、そのエントリの実行情報ID801と実行情報内部通番806を用いてB-Tre索引プリフェッチ情報820を参照し、対応エントリを求める。ここで対応エントリが求まらない場合には、プリフェッチ不可能としてステップ2339に進む。求まった場合には、ステップ2312に進む。なお、複数のエントリが求まる可能性がある。

【0197】ステップ2312では、読み出しデータからプリフェッチ先を求める処理を行う。ステップ2311において求めたB-Tre索引プリフェッチ情報820の全てのエントリにおける索引アクセス条件821を求め、アクセス先のデータの解釈をして索引アクセス条件821によりアクセスされるデータのボリューム名501とボリューム論理ブロック番号512を求める。このアクセス先は複数存在する可能性がある。

【0198】次アクセス先のボリューム名501とボリ

ューム論理ブロック番号512からキャッシュセグメント情報720を参照し、次データがデータキャッシュ28上に存在するかどうか確認する。次データが存在する場合には、再帰的に解釈処理を続ける。アクセス先が複数存在する場合には、それぞれで次アクセス先データのデータキャッシュ28上の存在確認とデータ解釈処理を実施する。データ解釈可能範囲ですべてのデータがデータキャッシュ28上に存在する場合にはプリフェッチ先なしとしてステップ2339に進み、そうでない場合には2313に進む。

【0199】ステップ2313では、現在確認中の実行情報のプリフェッチ用キャッシュのプールの残量を確認する。ステップ2311で求めた実行情報ID801を用いて実行情報ID管理情報800を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するDBMSキャッシュ管理ID811を求め、DBMSデータキャッシュ管理情報830を参照して、プリフェッチ用キャッシュのプールを管理するリストを取得し、その中の管理情報である領域セグメント数833を求める。その値が0でなかったら2315に進み、その値が0であったらステップ2339に進む。

【0200】ステップ2314では、ステップ2312で求めたプリフェッチ先に対してプリフェッチ用の領域を割り当て、プリフェッチコマンドを発行する。まず、ステップ2313で求めたプリフェッチ用キャッシュのプールを管理するリストからプリフェッチ用に割り当てるキャッシュセグメントを確保する。確保する量はステップ2312により求めたアクセス先に対応する量である。領域が不足する場合には可能な量全てを確保する。

【0201】この確保したキャッシュセグメントをB-Tre索引プリフェッチ情報820の対応するDBMSデータキャッシュ管理ID811で示されるDBMSデータキャッシュ管理情報830のエントリで管理されるプリフェッチデータ管理リストに繋ぎ、DBMSデータキャッシュ管理情報830のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報720中の対応する部分を更新する。確保したキャッシュセグメントにプリフェッチ先を割り当て、キャッシュセグメント情報720中の対応するエントリのボリューム名501とボリューム論理ブロック番号512を割り当て先の領域を示すように更新し、ステータス情報722の値を“インバリッド”に設定する。

【0202】このとき、キャッシュセグメントが要求量全てを確保できなかったときには、データ解釈段数を距離として、データ解釈を開始したデータに近いものに優先してプリフェッチ用キャッシュを割り当てる。アクセス先に関しては、ボリューム名501とボリューム論理ブロック番号512が求まっているため、データ構造物

理記憶位置情報712を参照して、アクセス先の物理記憶装置18上の領域である物理記憶装置名502と物理ブロック番号514とこのデータが属するDBMS名631とデータ構造名561を求める。

【0203】求めたDBMS名631とデータ構造名561を用いてDBMSデータ構造処理優先情報840を参照し、実効処理優先度842を求め、このアクセスの処理優先度とする。割り当てたキャッシュセグメントを用いてプリフェッチアクセス要求を作成する。その中には、物理記憶装置名502と物理ブロック番号514とデータを読み出し先のキャッシュのセグメントID721を含める。

【0204】求めたアクセスの処理優先度を処理優先度841とし指示子867に“プリフェッチ”を指定して、アクセス先の物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置18へのアクセス発行制御については後述する。その後、ステップ2339に進む。

【0205】ステップ2339においては、現在確認中のアクセス先データを保持するキャッシュセグメントがプリフェッチデータ管理リストに存在するか確認する。データを保持するキャッシュのセグメントID721を用いてキャッシュセグメント情報720を参照し、対応するリスト情報723を参照してどの管理リストにより管理されているかを判断する。プリフェッチデータの管理リストに存在する場合にはステップ2340に進み、そうでない場合にはステップ2341に進む。

【0206】ステップ2340においては、現在確認中のアクセス先データを保持するキャッシュセグメントを、そのキャッシュセグメントを管理するプリフェッチデータ管理リストを含む実行情報におけるプリフェッチ用キャッシュのプールに戻す処理を行う。ステップ2339で対応するプリフェッチデータ管理リストが求まっているため、それに対応するDBMSキャッシュ管理ID811を求める。この値をもとに、プリフェッチデータ管理リストを保持しているシーケンシャルプリフェッチ情報810またはB-Treepリフェッチ情報820のエントリを求める。どちらにも属していない場合には、既にプリフェッチ用キャッシュのプールに存在しているので以下のリストの変更処理は行わない。

【0207】求めたシーケンシャルプリフェッチ情報810またはB-Treepリフェッチ情報820のエントリにおける実行情報ID801を用いて実行情報ID管理情報800を参照して、キャッシュセグメントに戻すプリフェッチ用キャッシュのプールのリストを管理するDBMSキャッシュ管理ID811を求める。現在確認中のアクセス先データを保持するキャッシュセグメントを現在管理するプリフェッチデータ管理リストから外し、それをプリフェッチ用キャッシュのプールのリス

トに繋ぎ、それに対応してDBMSデータキャッシュ管理情報830のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報720中の対応する部分を更新する。これらの処理が完了後、ステップ2342に進み処理を完了する。

【0208】ステップ2341においては、現在確認中のアクセス先データを保持するキャッシュセグメント現在の管理リストから一旦外してそれをメインLRUリストのMRU側に接続し、対応するのキャッシュセグメント情報720のエントリとメインLRUリスト情報741と必要であれば再利用LRUリスト情報743をそれにあわせて更新する。この処理が完了後、ステップ2342に進み処理を完了する。

【0209】図20にディスクI/O実行管理情報860を利用した物理記憶装置18へのアクセス処理を行うバックグラウンド処理の処理フローを示す。この処理は、物理記憶装置18毎に実行され、記憶装置10が動作を開始すると同時にディスクコントローラ制御部52で開始され、その処理は無限ループとなっている。ステップ2401で処理を開始する。このとき、対象なる物理記憶装置18の識別子である物理記憶装置名502が指定される。

【0210】ステップ2402では、ディスクI/O実行管理情報860の対象とする物理記憶装置名502に対するI/O実行管理情報861中の処理優先度別処理残数カウンタ866の中の処理残りカウント値868を初期化する。ここでは、処理優先度情報850を参照し、処理優先度841毎の割り当てI/O比率851を求め、それにあらかじめ与えられた比率を乗じたものを処理優先度別処理残数カウンタ866の中の処理残りカウント値868の対応する部分に代入する。初期化後、ステップ2403に進む。

【0211】ステップ2403では、キャッシュセグメント利用管理情報740中のダーティセグメント数カウンタ746の値を確認する。その値があらかじめ定められたある閾値以上である場合には、ステップ2404に進み、そうでない場合には、ステップ2407に進む。

【0212】ステップ2404では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の書き込みキュー864に書き込み要求が存在するか確認する。存在する場合には、ステップ2405に進み、存在しない場合にはステップ2407に進む。

【0213】ステップ2405では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の書き込みキュー864の先頭に存在する書き込み要求を1つ取り出し、対象とする物理記憶装置名502を持つ物理記憶装置18への書き込み処理を実行する。その書き込み処理の完了を待ち、処理完了後、キャッシュセグメント情報720の更新を行う。対象とするセグメントID721はアクセス要求中に含まれているので、キ

キャッシュセグメント情報720中のそれに対応するエントリのステータス情報722を“ノーマル”に設定する。その後、ステップ2403に進む。なお、ここで、複数の書き込み要求を同時に実行してもよい。

【0214】ステップ2407では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中のホスト要求読出アクセスキュー中にアクセス要求が存在するか確認する。このとき、処理優先度別処理残数カウンタ866を考慮する。まず処理優先度別処理残数カウンタ866中の処理残りカウント値が0でない最も高い処理優先度841を求める。求めた処理優先度841と指示子867“ホスト要求読出”を用いて読み出しキュー配列863を参照し、対応するアクセスキューにアクセス要求が存在したらその処理優先度841を記憶してステップ2411に進む。存在しない場合、処理優先度別処理残数カウンタ866中の処理残りカウント値が0でない2番目に高い処理優先度841を求め、同じようにアクセス要求を確認し、存在したらその処理優先度を記憶してステップ2411に進む。

【0215】以下同様に処理優先度別処理残数カウンタ866中の処理残りカウント値を考慮しながら処理優先度841とその値におけるアクセス要求が存在するか確認する処理を繰り返し、存在する場合にはそのときの処理優先度を記憶してステップ2411に進み、存在しない場合にはステップ2408に進む。

【0216】ステップ2408では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中のプリフェッチアクセスキュー中にアクセス要求が存在するか確認する。このとき、処理優先度別処理残数カウンタ866を考慮する。まず処理優先度別処理残数カウンタ866中の処理残りカウント値が0でない最も高い処理優先度841を求める。

【0217】求めた処理優先度841と指示子867“プリフェッチ”を用いて読み出しキュー配列863を参照し、対応するアクセスキューにアクセス要求が存在したらその処理優先度841を記憶してステップ2412に進む。存在しない場合、以下同様に処理優先度別処理残数カウンタ866中の処理残りカウント値を考慮しながら処理優先度841とその値におけるアクセス要求が存在するか確認する処理を繰り返し、存在する場合にはそのときの処理優先度を記憶してステップ2412に進み、存在しない場合にはステップ2409に進む。

【0218】ステップ2409では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の書き込みキュー864に書き込み要求が存在するか確認する。存在する場合には、ステップ2405に進み、存在しない場合にはステップ2410に進む。

【0219】ステップ2410に到着するときは、アクセス要求が全く存在しないときか、あるいは、処理優先度別処理残数カウンタ866を考慮するために存在する

アクセス要求を実行できなかったときである。ここでは、読み出しキュー配列863とアクセス要求が書き込みキュー864にアクセス要求が全く存在しないか確認する。1つでも存在する場合には、即座にステップ2402に進む。全く存在しない場合には、新たにアクセス要求が発行されるまでこのステップで待ち、アクセス要求が発行された時点で、ステップ2402に進む。

【0220】ステップ2411では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863をステップ2407で記憶した処理優先度841と指示子867として“ホスト要求読出”を参照し、対応するアクセスキューに存在するアクセス要求を1つ取り出し、対象とする物理記憶装置名502を持つ物理記憶装置18からの読み出し処理を実行する。

【0221】その読み出し処理の完了を待ち、処理完了後、キャッシュセグメント情報720の更新を行う。対象とするセグメントID721はアクセス要求中に含まれているので、キャッシュセグメント情報720中のそれに対応するエントリのステータス情報722を“ノーマル”に設定する。その後、ステップ2413に進む。

【0222】なお、ここで、複数の読み出し要求を同時に実行してもよい。ただし、その上限値は処理優先度別処理残数カウンタ866のステップ2407で記憶された処理優先度841に対応する処理残りカウント値868の値である。

【0223】ステップ2412では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の読み出しキュー配列863をステップ2408で記憶された処理優先度841と指示子867として“プリフェッチ”を参照し、対応するアクセスキューに存在するアクセス要求を1つ取り出し、対象とする物理記憶装置名502を持つ物理記憶装置18からの読み出し処理を実行する。

【0224】その読み出し処理の完了を待ち、処理完了後、キャッシュセグメント情報720の更新を行う。対象とするセグメントID721はアクセス要求中に含まれているので、キャッシュセグメント情報720中のそれに対応するエントリのステータス情報722を“ノーマル”に設定する。

【0225】その後、ステップ2413に進む。なお、ここで、複数の読み出し要求を同時に実行してもよい。ただし、その上限値は処理優先度別処理残数カウンタ866のステップ2408で記憶された処理優先度841に対応する処理残りカウント値868の値である。

【0226】ステップ2413では、対象とする物理記憶装置名502に対応するI/O実行管理情報861中の処理優先度別処理残数カウンタ866の処理残りカウント値868の更新を行う。ステップ2411もしくは2412で実行した読み出しアクセス数を、その読み出

し要求が存在していた処理優先度 841 に対応する処理残りカウンタ値 868 から減ずる。

【0227】ステップ 2414 では、対象とする物理記憶装置名 502 に対応する I/O 実行管理情報 861 中の処理優先度別処理残数カウンタ 866 の処理残りカウンタ値 868 の確認を行う。対象である処理優先度別処理残数カウンタ 866 中の全ての処理残りカウンタ値 868 の値が 0 以下の場合には、ステップ 2402 に進み、そうでない場合にはステップ 2403 に進む。

【0228】図 21 にキャッシュ制御部 54 がバックグラウンドで実行する周期処理の処理フローを示す。この処理は、ダーティセグメント数を一定値以下に、再利用 LRU リストに存在するセグメント数を一定値以上に保つための処理で、記憶装置 10 が動作を開始すると同時に例えば 1 秒周期等の一定周期で処理が開始される。ステップ 2501 で処理が開始される。

【0229】ステップ 2502 では、キャッシュセグメント利用管理情報 740 中のダーティセグメント数カウンタ 746 の値を確認する。その値があらかじめ定められた閾値以上の場合にはステップ 2503 に進み、閾値未満の場合にはステップ 2504 に進む。

【0230】ステップ 2503 では、ダーティセグメント数カウンタ 746 の値が前記閾値未満になるようにキャッシュセグメント情報 720 中のステータス情報 722 が“ダーティ”状態のキャッシュセグメントの一部のデータを物理記憶装置 18 に対して書き込むアクセス要求を発行する処理を行う。ここで、書き込むセグメント数は、ダーティセグメント数カウンタ 746 から前記閾値を引いた値にあらかじめ定められている一定値を加えたものとする。

【0231】メイン LRU リストの LRU 側からキャッシュセグメント情報 720 中のステータス情報 722 が“ダーティ”であるキャッシュセグメントを先に求めた書き込みセグメント数だけ選択し、そのステータス情報 722 を“ライト”に設定する。そして、求めたキャッシュセグメントのセグメント ID 721 とそのセグメントに記憶されているデータのボリューム名 501 とボリューム論理ブロック番号 512 を求める。

【0232】求めたボリューム名 501 とボリューム論理ブロック番号 512 から対応する物理記憶装置名 502 と物理論理番号 514 を求め、更にそのデータを保持するキャッシュのセグメント ID 721 を指定して書き込みアクセス要求を作成し、ディスク I/O 実行管理情報 860 中のアクセス先の物理記憶装置名 502 に対応する I/O 実行管理情報 861 中の書き込みキュー 864 中に追加する。その後、作成した書き込みアクセス要求の数だけダーティセグメント数カウンタ 746 の値を減じ、ステップ 2504 に進む。

【0233】ステップ 2504 では、キャッシュセグメント利用管理情報 740 中の再利用 LRU リスト情報 7

43 中に記憶されているそのリストに存在するセグメント数を確認する。その値があらかじめ定められた閾値未満の場合にはステップ 2505 に進み、閾値以上の場合にはステップ 2506 に進み、処理を完了する。

【0234】ステップ 2505 では、メイン LRU リストに存在するセグメントの一部を再利用 LRU リストに移動し、再利用 LRU リストに存在するセグメント数を増加させる処理を行う。ここで、移動させるセグメント数は、再利用 LRU リスト情報 743 中のリストに存在するセグメント数から前記閾値を引いた値にあらかじめ定められている一定値を加えたものとする。

【0235】メイン LRU リストの LRU 側からキャッシュセグメント情報 720 中のステータス情報 722 が“ノーマル”であるキャッシュセグメントを先に求めた移動させるセグメント数だけ選択してメイン LRU リストから外し、それらを再利用 LRU リストの MRU 側に繋ぎ直す。そして、対応するキャッシュセグメント情報 720 のエントリとメイン LRU リスト情報 741 と再利用 LRU リスト情報 743 をそれにあわせて更新する。その後、ステップ 2506 に進み、処理を完了する。＜第二の実施の形態＞本実施の形態では、DBMS が実行される計算機と、ファイルを管理単位とし、データキャッシュを保持する記憶装置が接続された計算機システムにおいて、記憶装置が DBMS に関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、DBMS で実行されるクエリの実行プラン情報、DB の処理優先順位情報を取得し、それらを用いて記憶装置がより好ましいアクセス性能を提供する。記憶装置は、DBMS に関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、DBMS で実行されるクエリの実行プランを利用することにより、DBMS がこれからどのデータをどの順序でどのようにアクセスするかを把握することができる。

【0236】そこで、この把握したアクセス方法に関する情報を利用して、あらかじめ利用される可能性が高いデータを記憶装置上のデータキャッシュ上に用意しておくことにより、DBMS に対してより高いアクセス性能を提供する。また、DB の処理優先順位情報を利用し、処理優先度が高い DB のデータに対して、記憶装置が保持する物理記憶装置へアクセスを優先的に実施したり、また、データキャッシュの利用量をより多く割り当てたりすることにより、処理優先度が高い DB のデータに対するアクセス性能を向上させる。

【0237】図 22 は、本発明の第二の実施の形態における計算機システムの構成図である。図示されたように、本発明の第二の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【0238】本実施の形態においては I/O パスインターフェイス 70、I/O パス 71、I/O パススイッチ 72 が存在せず、記憶制御装置 10b と DB ホスト 80

c, 80dはネットワーク79を介してのみ接続される。記憶装置10はファイルを単位としたデータ記憶管理を行う記憶装置10bに変更される。そのため、物理記憶装置稼動情報32、データキャッシュ管理情報34、処理優先度付ディスクI/O管理情報36、DBMS実行情報38、DBMSデータ情報40、ボリューム物理記憶位置管理情報42がそれぞれ物理記憶装置稼動情報32b、データキャッシュ管理情報34b、処理優先度付ディスクI/O管理情報36b、DBMS実行情報38b、DBMSデータ情報40b、ファイル記憶管理情報42bに変更される。

【0239】DBホスト80c, 80dで実行されるOS100ではボリュームマネージャ102、ファイルシステム104が削除されその代わりに記憶装置10bが提供するファイルをアクセスするための機能を有するネットワークファイルシステム104bが追加され、OS100が保持するマッピング情報106がネットワークマウント情報106bへ変更される。

【0240】記憶装置10はファイルを管理単位とする記憶装置10bに変更される。DBホスト80c, 80dからのアクセスもNFS等のファイルをベースとしたプロトコルで実施される。記憶装置10におけるボリュームの役割は、記憶装置10bにおいてはファイルもしくはファイルを管理するファイルシステムとなり、そのファイルの記憶位置管理情報がファイル記憶管理情報42bである。1つの記憶装置10bの中に複数のファイルシステムが存在しても構わない。物理記憶装置18の稼動情報はボリュームを単位とした取得からファイルシステムまたはファイルを単位とした取得に変更する。記憶装置10b内にファイルシステムが存在する場合でもデータの移動機能を実現可能である。

【0241】図23はDBホスト80c, 80dのOS100内に記憶されているネットワークマウント情報106bを示す。ネットワークマウント情報106bは、記憶装置10bから提供され、DBホスト80c, 80dにおいてマウントされているファイルシステムの情報で、ファイルシステムの提供元記憶装置とそのファイルシステムの識別子である記憶装置名583とファイルシステム名1001、そして、そのファイルシステムのマウントポイントの情報であるマウントポイント1031の組を保持する。

【0242】図24は記憶装置10b内に保持されるファイル記憶管理情報42bを示す。図4のボリューム物理記憶位置管理情報42からの変更点は、ボリュームの識別子であるボリューム名501がファイルの識別子となるファイルシステム名1001とファイルパス名1002に、ボリューム内のデータ領域を示すボリューム論理ブロック番号512がファイルブロック番号1003変更されるものである。

【0243】図25に記憶装置10b内に保持される物

理記憶装置稼動情報32bを示す。図5の物理記憶装置稼動情報32からの変更点は、稼動情報取得単位がボリュームからファイルシステムに変更されたため、ボリューム名501の部分がファイルシステム名1001に変更されたことである。また、稼動情報取得単位をファイルとしてもよく、このときはボリューム名501の部分がファイルシステム名1001とファイルパス名1002に変更される。図26に記憶装置10b内に保持されているDBMSデータ情報40bを示す。図6のDBMSデータ情報40からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたためデータ構造物理記憶位置情報712に修正が加えられ、データ構造物理記憶位置情報712bに変更されたことと、DBMSスキーマ情報711中のDBMSホストマッピング情報627中のマッピング情報648が保持するデータがDBホストにおけるマッピング情報106からネットワークマウント情報1066に変更されたことである。

【0244】図27にDBMSデータ情報40b中に含まれるデータ構造物理記憶位置情報712bを示す。図8のデータ構造物理記憶位置情報712からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、ボリューム名501とボリューム論理ブロック番号512の部分がファイルシステム名1001とファイルパス名1002とファイルブロック番号1003に変更されたことである。この情報は、DBMSスキーマ情報711内のDBMSデータ記憶位置情報622とDBMSホストマッピング情報627とファイル記憶管理情報42bを参照して、対応する部分を組み合わせることにより作成する。

【0245】図28に記憶装置10b内に保持されているデータキャッシュ管理情報34bを示す。図10のデータキャッシュ管理情報34からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、キャッシュセグメント情報720に修正が加えられ、キャッシュセグメント情報720bに変更されたことである。キャッシュセグメント情報720bのキャッシュセグメント情報720からの変更点は、上述の理由により、ボリューム名501とボリューム論理ブロック番号512の部分がファイルシステム名1001とファイルパス名1002とファイルブロック番号1003に変更されたことである。

【0246】記憶装置10bにおける本実施の形態における第一の実施の形態からの差は、ほとんどがボリューム名501をファイルシステム名1001とファイルパス名1002に、ボリューム論理ブロック番号512をファイルブロック番号1003に変更することであり、その他の変更点もその差を述べてきた。記憶装置10bにおける処理に関しても基本的にこれまで述べてきた変更点と同じ変更点への対応方法を実施すれば、第一の実

施の形態における処理をほぼそのまま本実施の形態に当てはめることができる。

【0247】

【発明の効果】本発明により以下のことが可能となる。第一に、DBMSが管理するデータを保持する記憶装置において、DBMS向けのアクセスの最適化が実現される。この記憶装置を用いることにより、既存のDBMSに対してプログラムの修正無しにDBMS稼動システムの性能を向上させることができる。つまり、高性能なDBシステムを容易に構築できる。

【0248】第二に、DBMSが管理するデータを保持する記憶装置において、DBのデータや処理に与えられた処理優先度を考慮するアクセスの最適化が実現される。DBのデータ毎の処理優先度や処理毎の優先度を考慮することにより、特定のDBに対する処理性能を保持するDBシステムが実現できる。つまり、DBの処理性能に対するSLA (Service Level Agreement) を守るシステムを容易に実現できる。また、特定のDBの処理性能が保持されるシステムを構築できるため、DBシステムの性能に関する管理コストを削減できる。

【図面の簡単な説明】

【図1】第一の実施の形態における計算機システムの構成を示す図である。

【図2】DBホスト80a、80bのOS100内に記憶されているマッピング情報106を示す図である。

【図3】DBMS110a、110b内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報114を示す図である。

【図4】記憶装置10内に保持されているボリューム物理記憶位置管理情報42を示す図である。

【図5】記憶装置10内に保持されている物理記憶装置稼動情報32を示す図である。

【図6】記憶装置10内に保持されているDBMSデータ情報40を示す図である。

【図7】DBMSデータ情報40中に含まれるDBMSスキーマ情報711を示す図である。

【図8】DBMSデータ情報40中に含まれるデータ構造物理記憶位置情報712を示す図である。

【図9】記憶装置10内に保持されているDBMS実行情報38を示す図である。

【図10】記憶装置10内に保持されているデータキャッシュ管理情報34を示す図である。

【図11】記憶装置10内に保持されている処理優先度付ディスクI/O管理情報36を示す図である。

【図12】クエリ871とその処理を実現するためにDBMS110aが作成したクエリ実行プラン872を図示したものである。

【図13】記憶装置10に与えられるクエリ実行プランに関する情報であるクエリプラン情報880を示す図で

ある。

【図14】記憶装置10がクエリプラン情報880を受け取った際の処理フローを示す図である。

【図15】記憶装置10がクエリプラン情報880に対応するクエリの完了通知を受け取った時の処理フローを示す図である。

【図16】記憶装置10がDBホスト80a、80bから書き込みアクセス要求を受け取ったときの処理フローを示す図である。

【図17】記憶装置10がDBホスト80a、80bから読み出しアクセス要求を受け取ったときの処理フローを示す図である。

【図18】DBデータの読み出しアクセス後の処理フローを示す図（その1）である。

【図19】DBデータの読み出しアクセス後の処理フローを示す図（その2）である。

【図20】ディスクI/O実行管理情報860を利用した物理記憶装置18へのアクセス処理を行うバックグラウンド処理の処理フローを示す図である。

【図21】バックグラウンドで実行するダーティセグメント数と再利用LRUリスト中存在セグメント数の管理のための周期処理の処理フローを示す図である。

【図22】第二の実施の形態における計算機システムの構成を示す図である。

【図23】DBホスト80c、80dのOS100内に記憶されているネットワークマウント情報106bを示す図である。

【図24】記憶装置10b内に保持されるファイル記憶管理情報42bを示す図である。

【図25】記憶装置10b内に保持される物理記憶装置稼動情報32bを示す図である。

【図26】記憶装置10b内に保持されているDBMSデータ情報40bを示す図である。

【図27】DBMSデータ情報40b中に含まれるデータ構造物理記憶位置情報712bを示す図である。

【図28】記憶装置10b内に保持されているデータキャッシュ管理情報34bを示す図である。

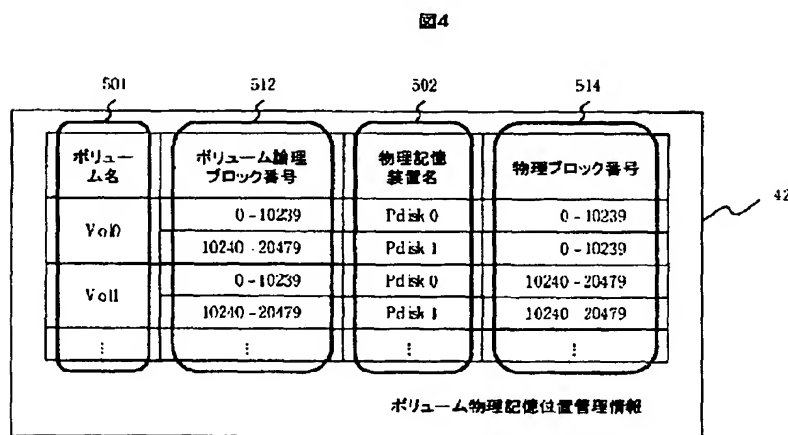
【符号の説明】

| | |
|---------|-------------------|
| 10, 10b | 記憶装置 |
| 18 | 物理記憶装置 |
| 28 | データキャッシュ |
| 32, 32b | 物理記憶装置稼動情報 |
| 34, 34b | データキャッシュ管理情報 |
| 36, 36b | 処理優先度付ディスクI/O管理情報 |
| 38 | DBMS実行情報 |
| 40, 40b | DBMSデータ情報 |
| 42 | ボリューム物理記憶位置管理情報 |

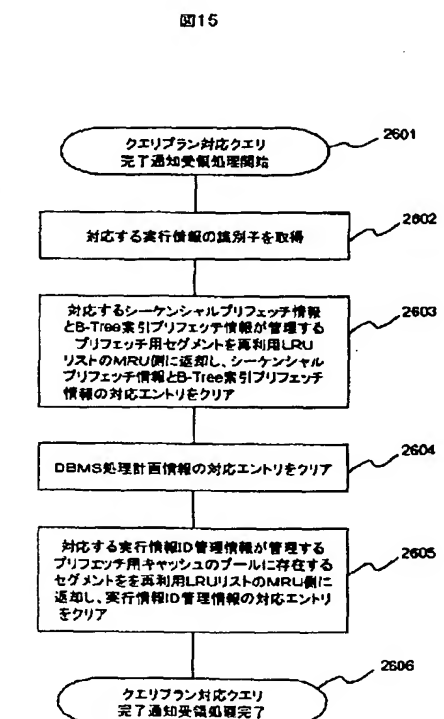
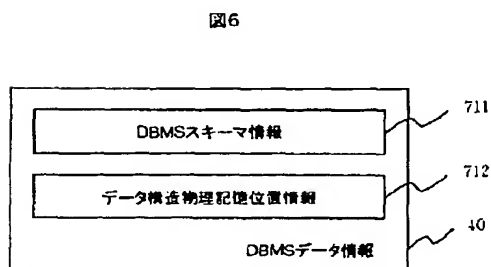
| | | | |
|------------------------|--------------|--------------|---------------|
| 42 b | ファイル記憶管理情報 | 100 | OS (オペレーティングシ |
| 50 | 記憶装置制御プログラム | ステム) | |
| 52 | ディスクコントローラ制御 | 102 | ボリュームマネージャ |
| 部 | | 104 | ファイルシステム |
| 54 | キャッシュ制御部 | 104 b | ネットワークファイルシス |
| 56 | 物理記憶位置管理部 | テム | |
| 58 | I/Oバスインターフェイ | 106 | マッピング情報 |
| ス制御部 | | 106 b | ネットワークマウント情報 |
| 60 | ネットワークインターフェ | 110 a, 110 b | DBMS (データベース管 |
| イス制御部 | | 理システム) | |
| 70 | I/Oバスインターフェイ | 114 | スキーマ情報 |
| ス | | 116 | DBMS情報通信部 |
| 71 | I/Oバス | 118 | DBMS情報取得・通信プ |
| 72 | I/Oバススイッチ | ログラム | |
| 78 | ネットワークインターフェ | 120 | クエリプラン取得プログラ |
| イス | | ム | |
| 79 | ネットワーク | 126 | DBMSフロントエンドプ |
| 80 a, 80 b, 80 c, 80 d | DBホスト | ログラム | |
| 81 | DBクライアント | 130 | ホスト情報設定プログラム |
| 82 | 処理性能管理サーバ | 132 | 処理性能管理プログラム |

【図4】

【図15】

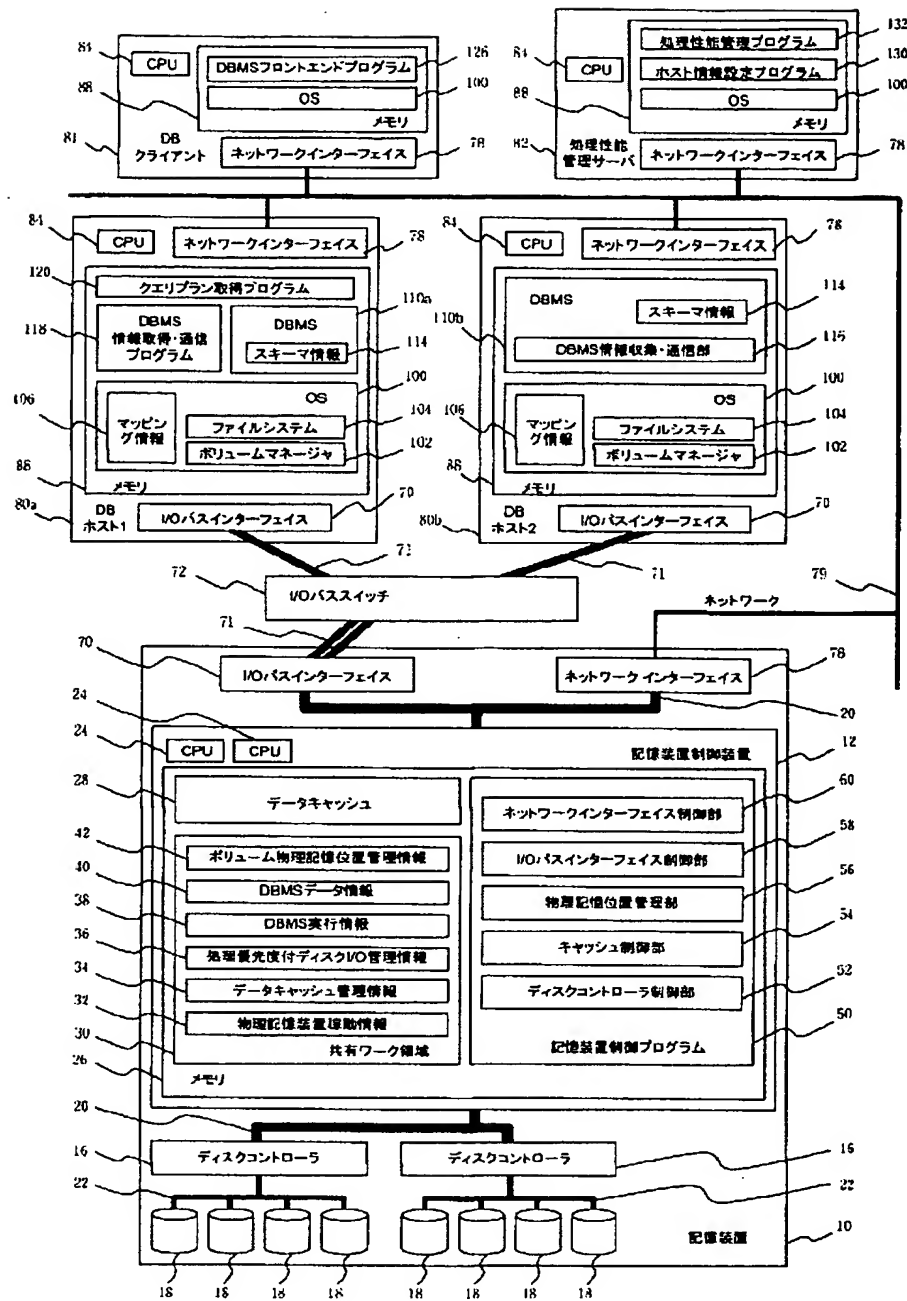


【図6】



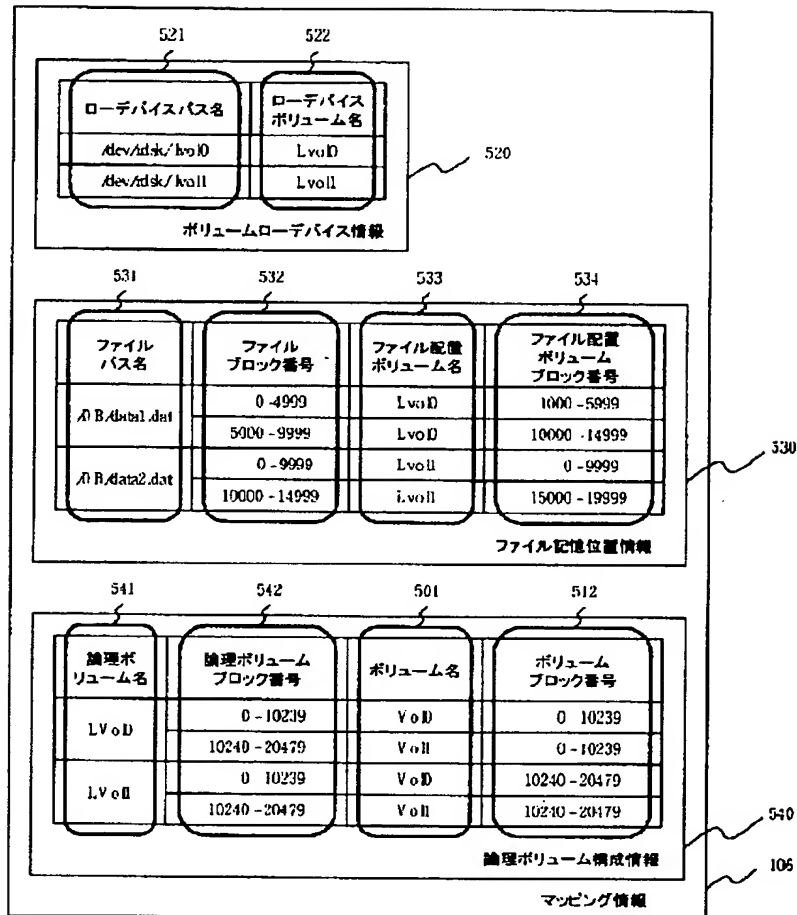
【図1】

図1



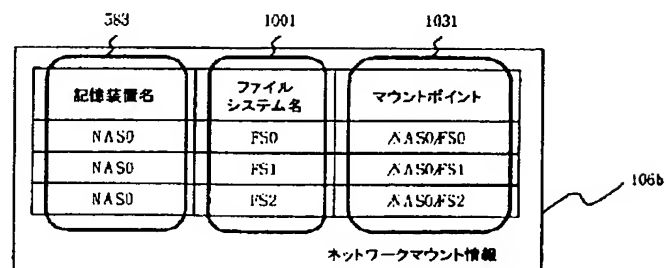
【図2】

図2



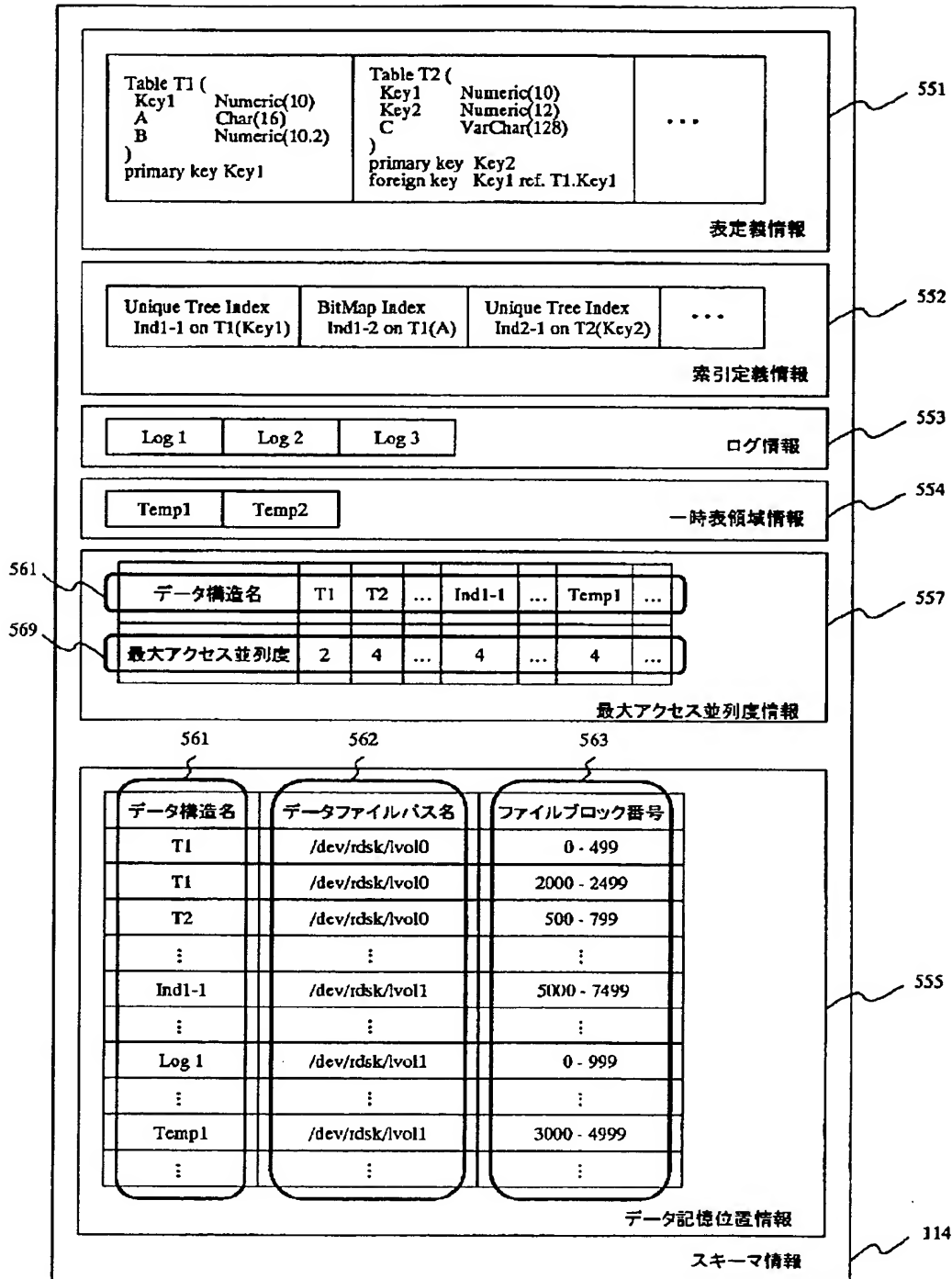
【図23】

図23



【図3】

図3



【图 2 6】

图 26

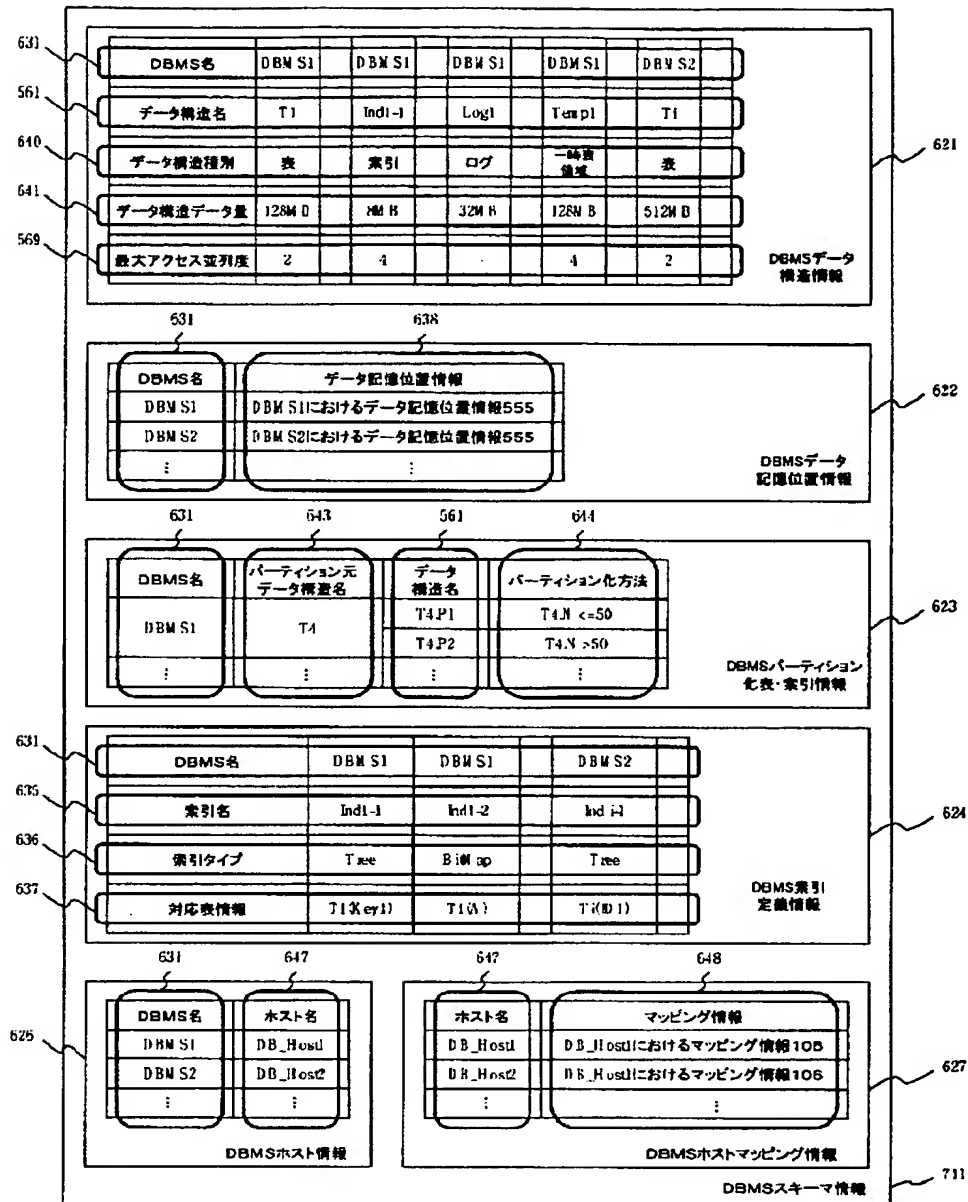
Figure 1 is a block diagram of a DBMS Data Information structure. It consists of a large rectangular container labeled "DBMSデータ情報" (DBMS Data Information) at the bottom right. Inside this container, there are two smaller rectangular blocks. The top block is labeled "DBMSスキーマ情報" (DBMS Schema Information) and is connected to the reference numeral 711. The bottom block is labeled "データ構造物理記憶位置情報" (Data Structure Physical Storage Location Information) and is connected to the reference numeral 712a. The reference numeral 40b is also shown, pointing to the overall container structure.

| | | | | | | | |
|-------------|--------|-------------|-------------|--|-------------|--|-----|
| DBMS名 | DBMS1 | DBMS1 | DBMS1 | | DBMS2 | | 631 |
| データ構造名 | T1 | T1 | T2 | | Ti | | 561 |
| データ構造ブロック番号 | 0-4999 | 5000-9999 | 0-239 | | 0-9999 | | 716 |
| ボリューム名 | V00 | V01 | V00 | | V02 | | 501 |
| ボリュームブロック番号 | 0-4999 | 5000-9999 | 10000-10239 | | 0-9999 | | 512 |
| 物理記憶装置名 | Pdisk0 | Pdisk0 | Pdisk0 | | Pdisk1 | | 502 |
| 物理ブロック番号 | 0-4999 | 10240-15239 | 10000-10239 | | 20480-30479 | | 514 |

データ構造物理記憶位置情報 712

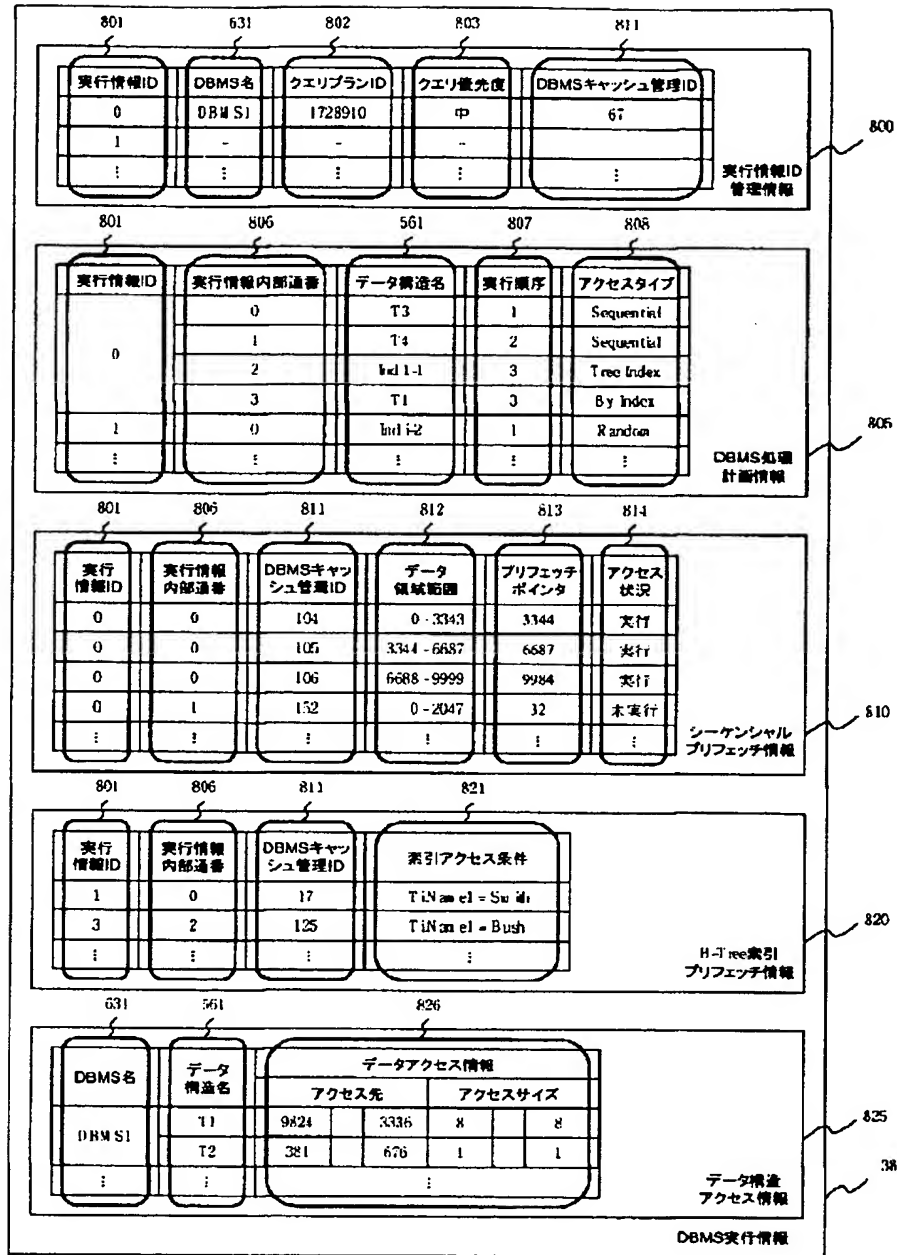
【図7】

図7



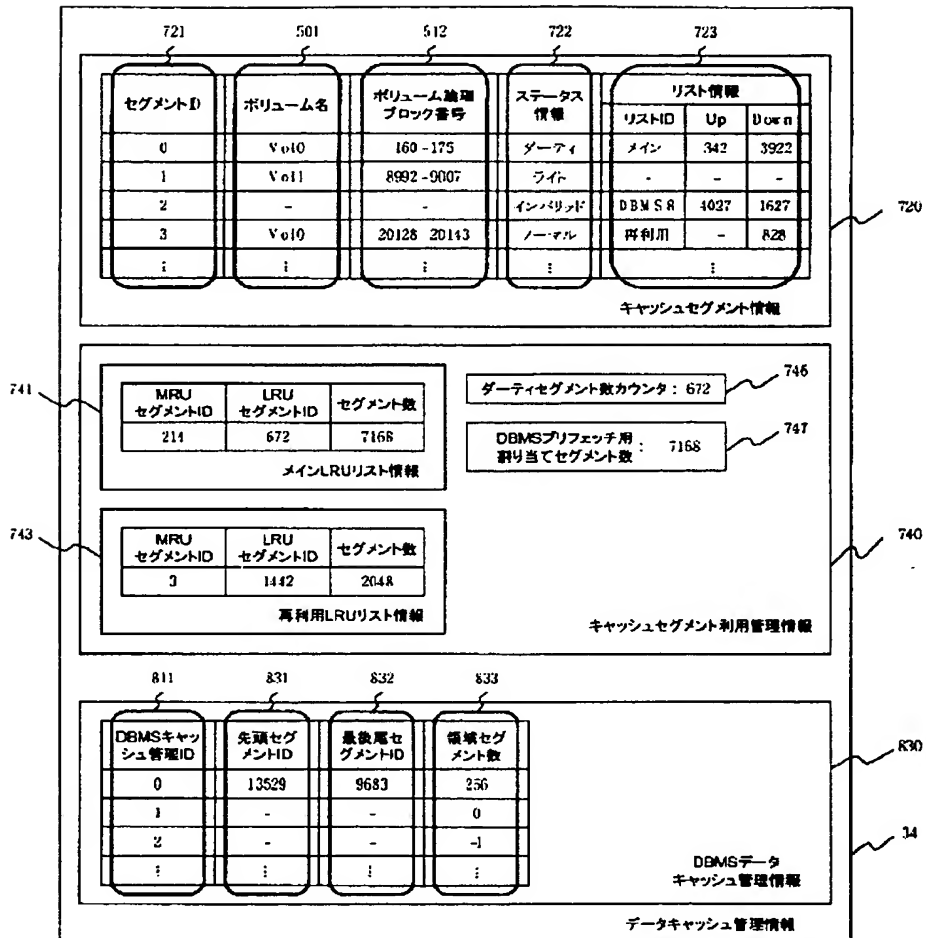
【図9】

図9



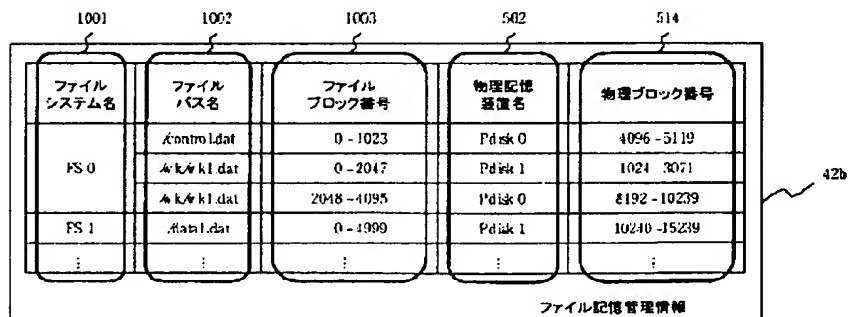
【図10】

図10



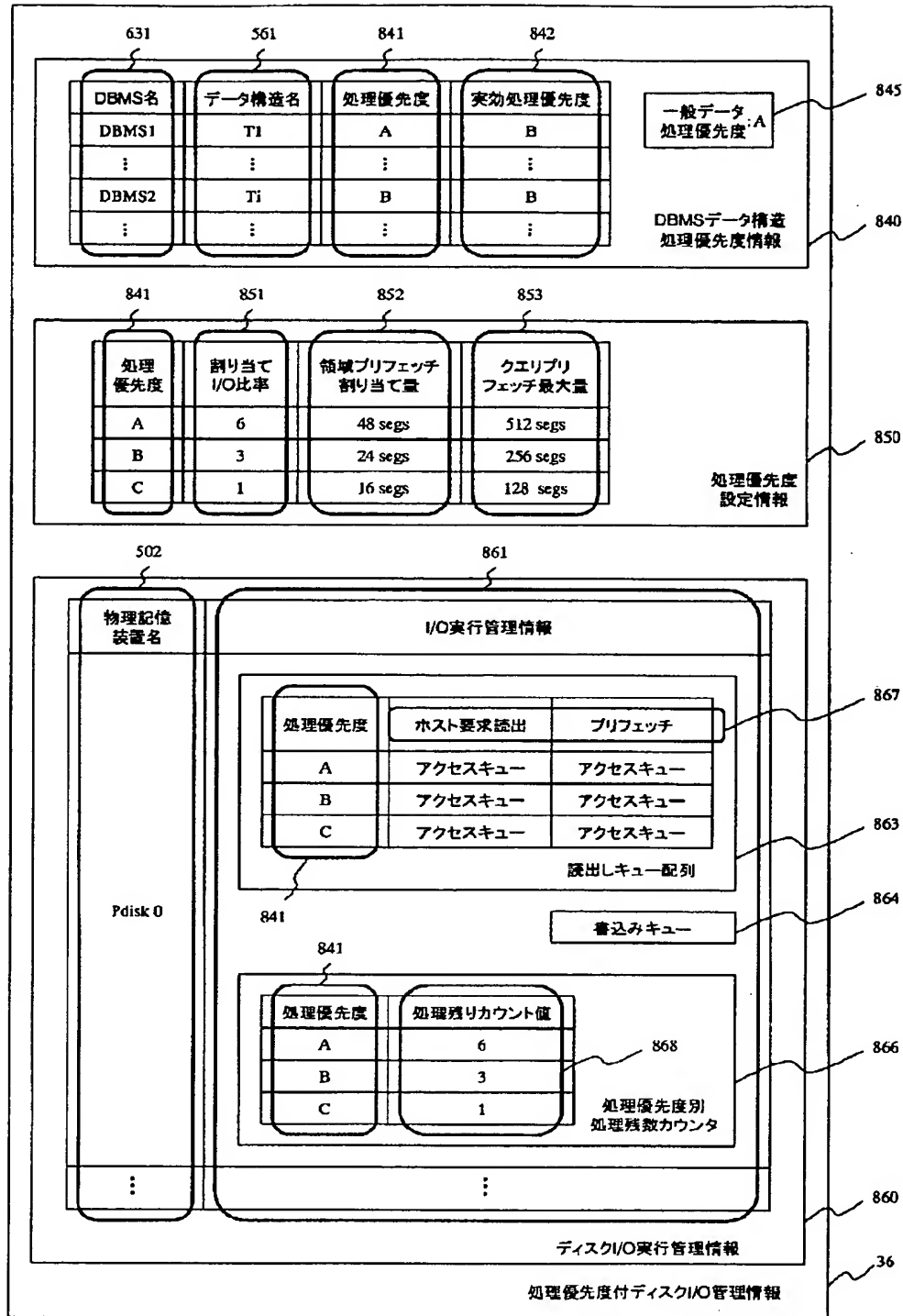
【図24】

図24



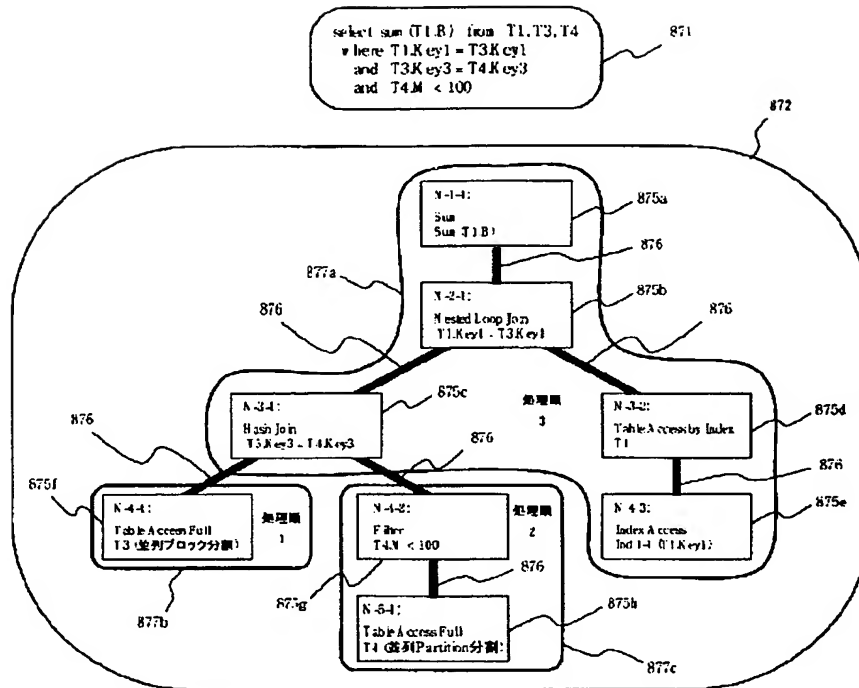
【図11】

図11



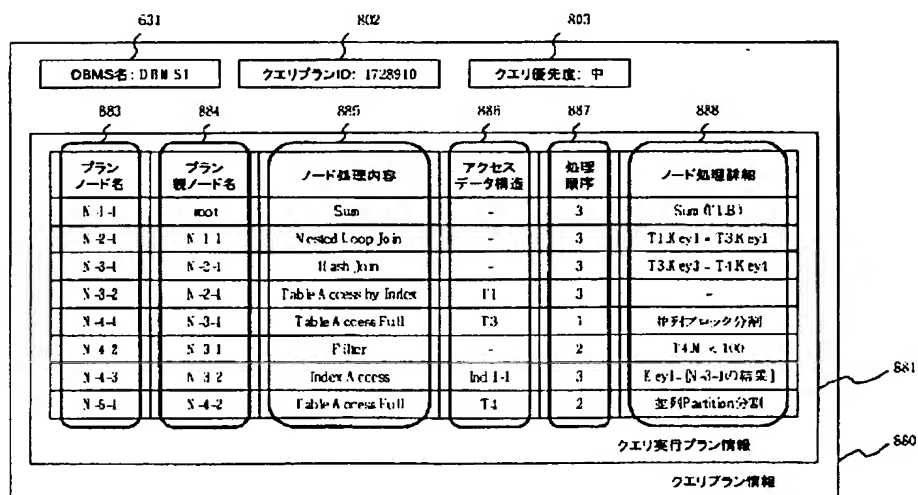
【図12】

図12

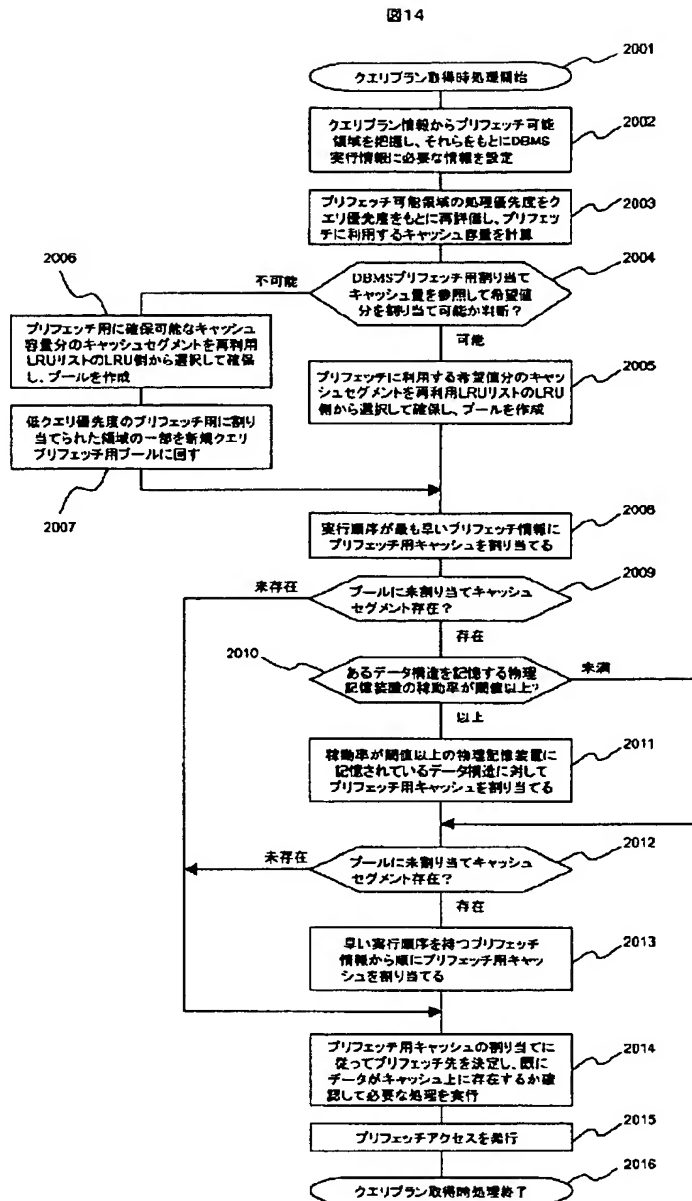


【図13】

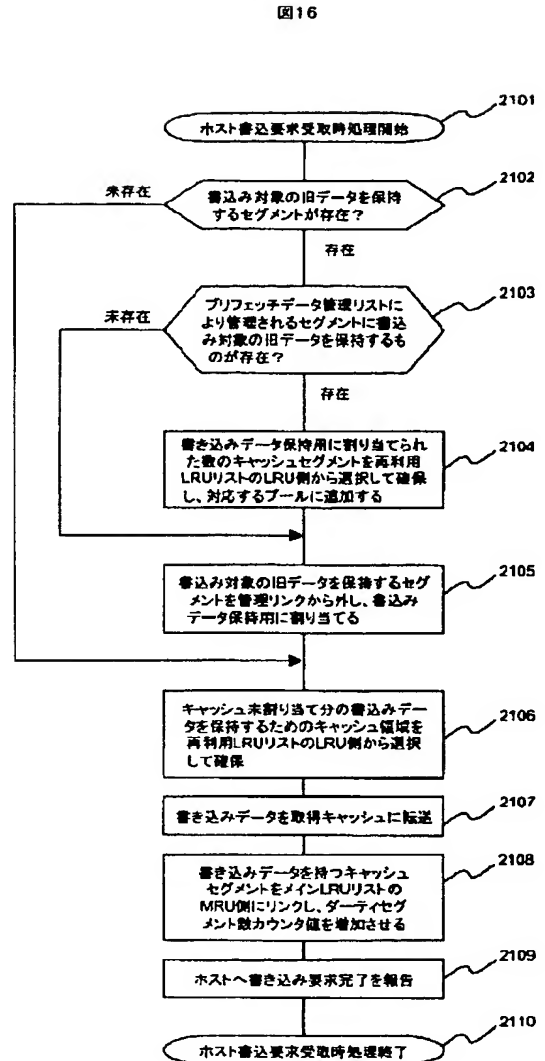
図13



【図14】

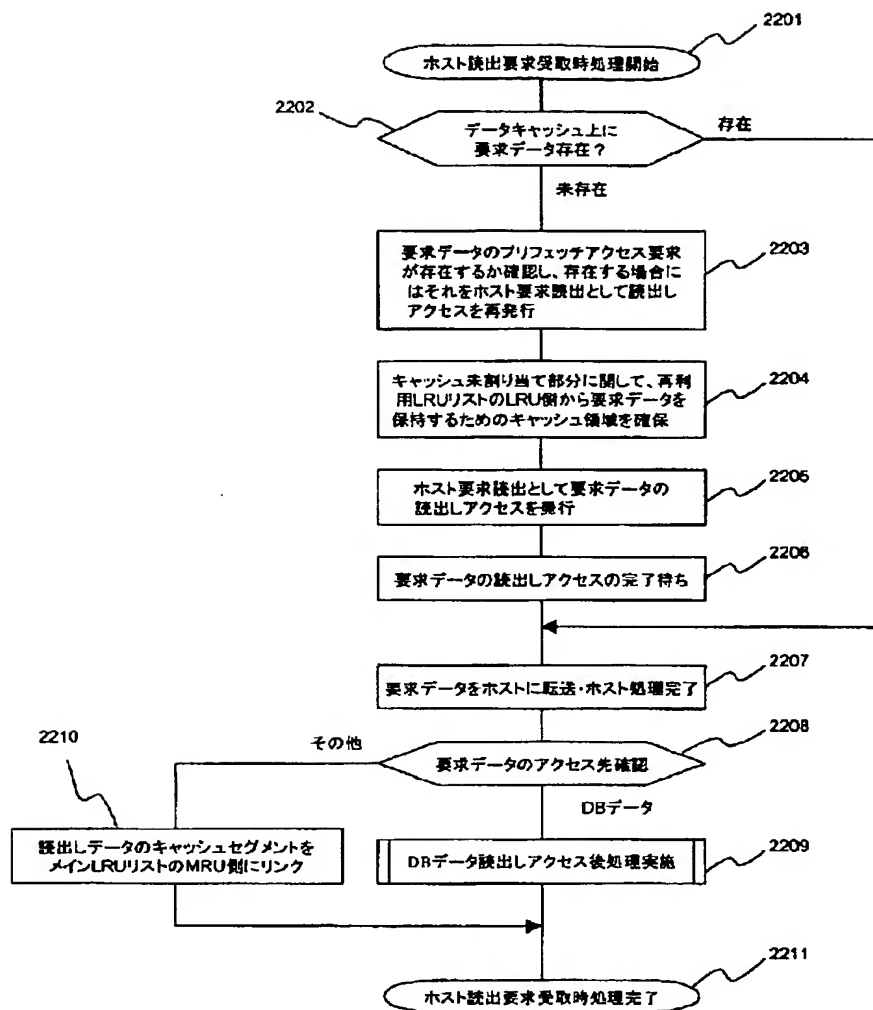


【図16】



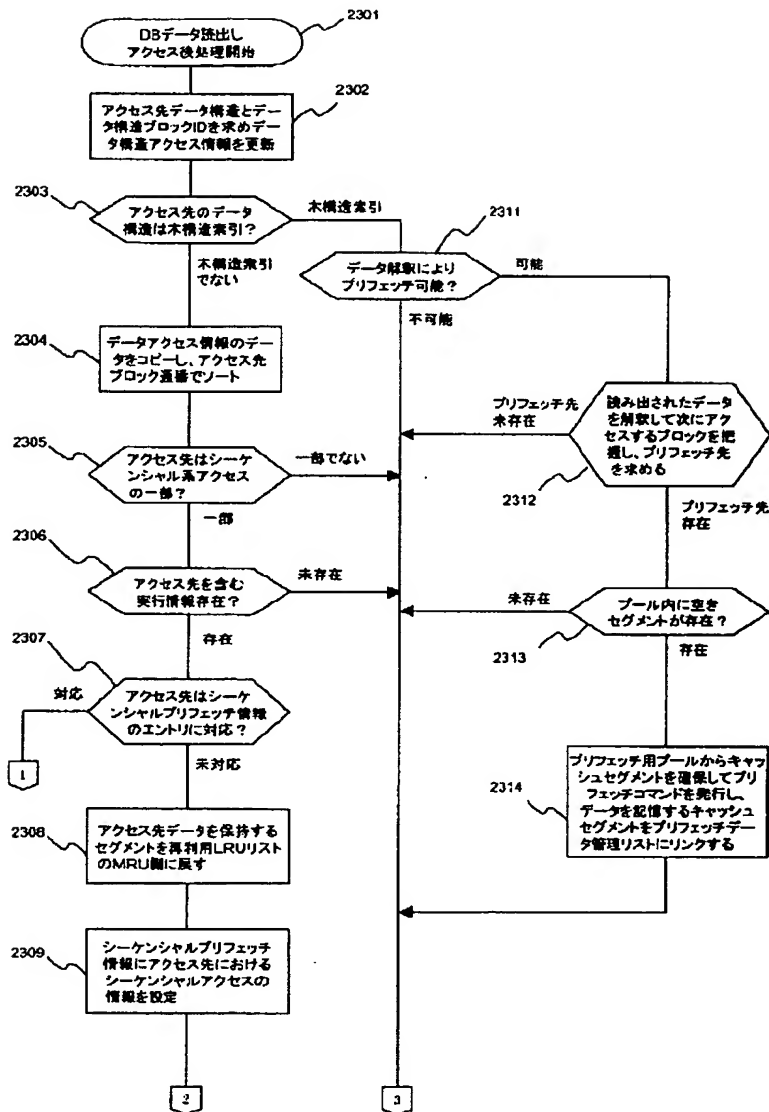
【図17】

図17



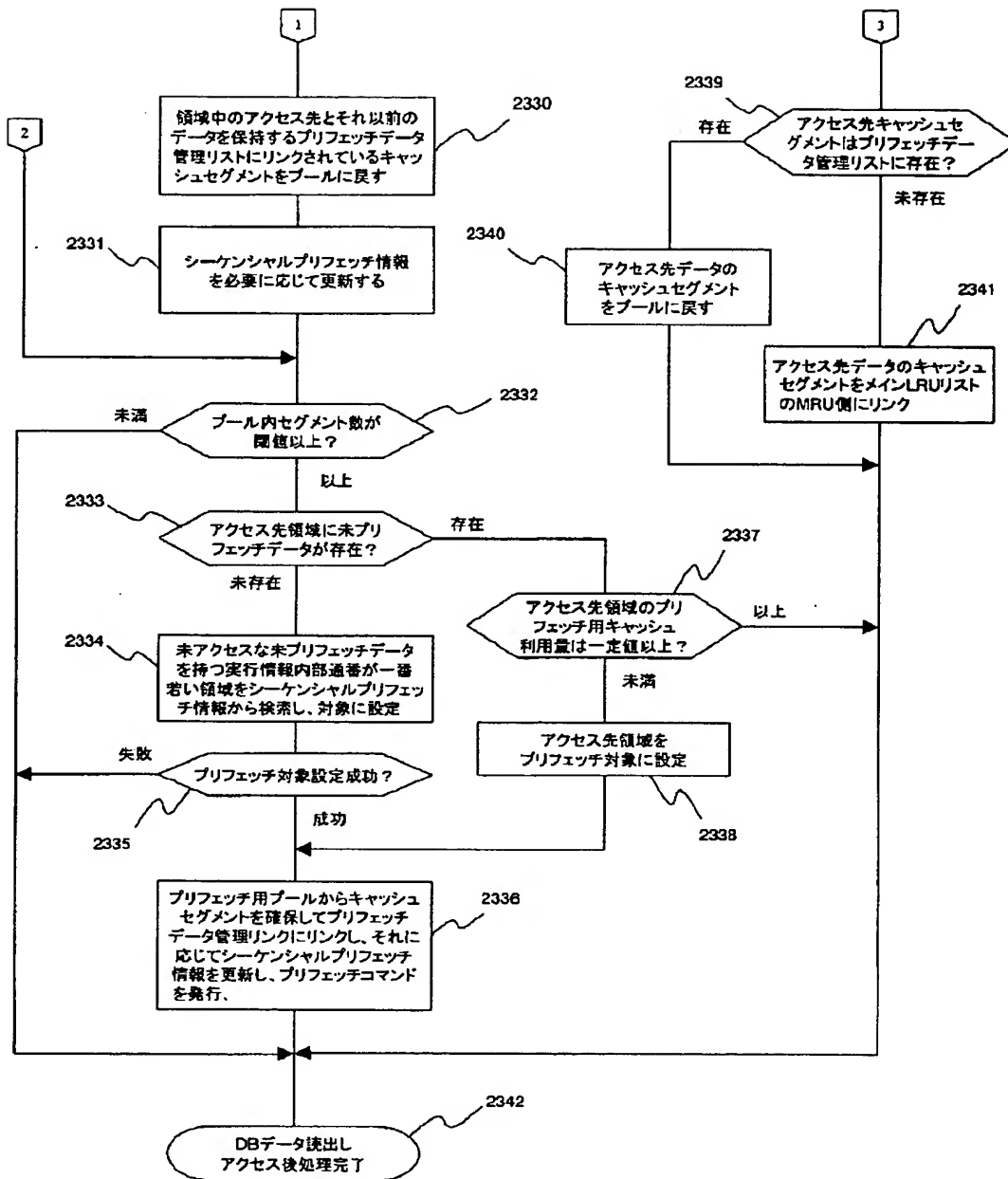
【図18】

図18



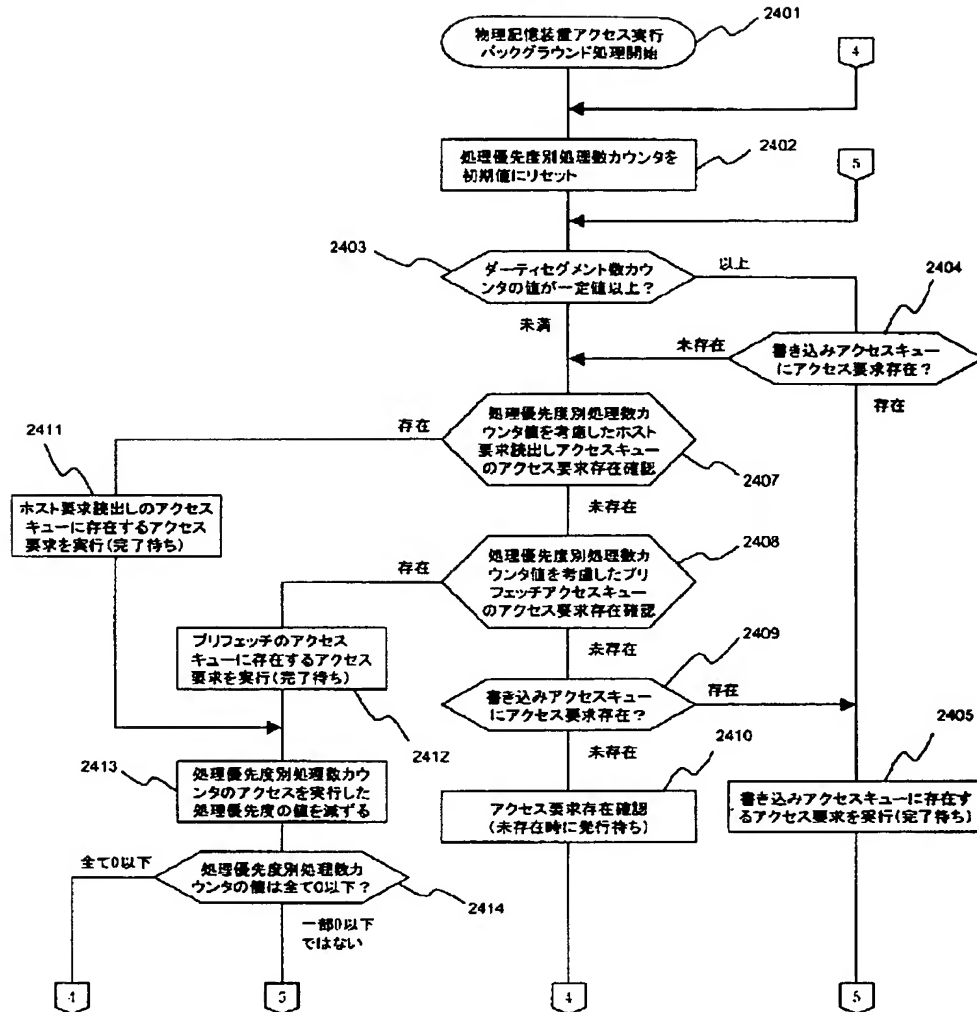
【図19】

図19



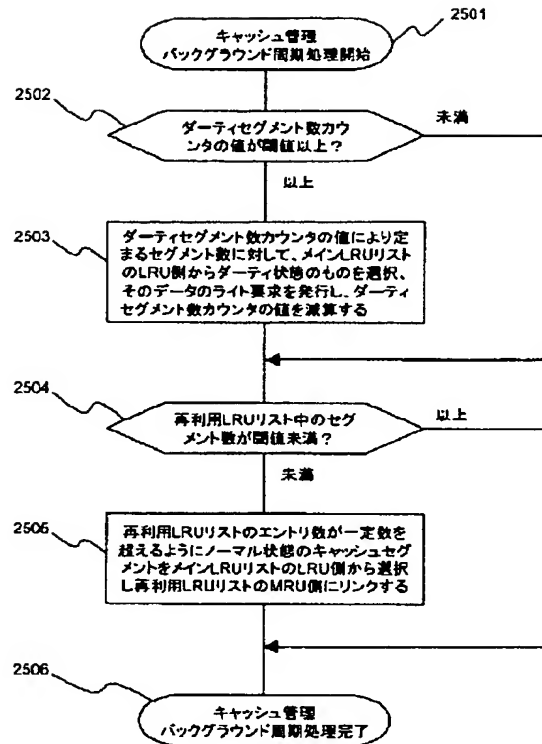
【図20】

図20



【図21】

図21



【図25】

図25

| | | | | | | |
|-------------------------------|---------------------------------|------------|----------|---------|---|------|
| ファイルシステム名 | | Vol0 | Vol0 | Vol1 | | 1001 |
| 物理記憶装置名 | | Pdisk 0 | Pdisk 1 | Pdisk 0 | | 502 |
| 累積稼働時間 | | 23917390 | 38902849 | 8012891 | | 592 |
| 旧累積稼働時間 | | 22787638 | 38783484 | 7592039 | | 593 |
| 稼働率 | 2000/4/1 12:00 ~ 2000/4/1 12:15 | 20% | 12% | 4% | | 594 |
| | 2000/4/1 12:15 ~ 2000/4/1 12:30 | 15% | 10% | 7% | | |
| | 2000/4/1 12:30 ~ 2000/4/1 12:45 | 16% | 9% | 5% | | |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |
| 前回累積稼働時間取得時刻: 2001/4/12 18:15 | | 物理記憶装置稼働情報 | | | | 32b |

【図22】

図22

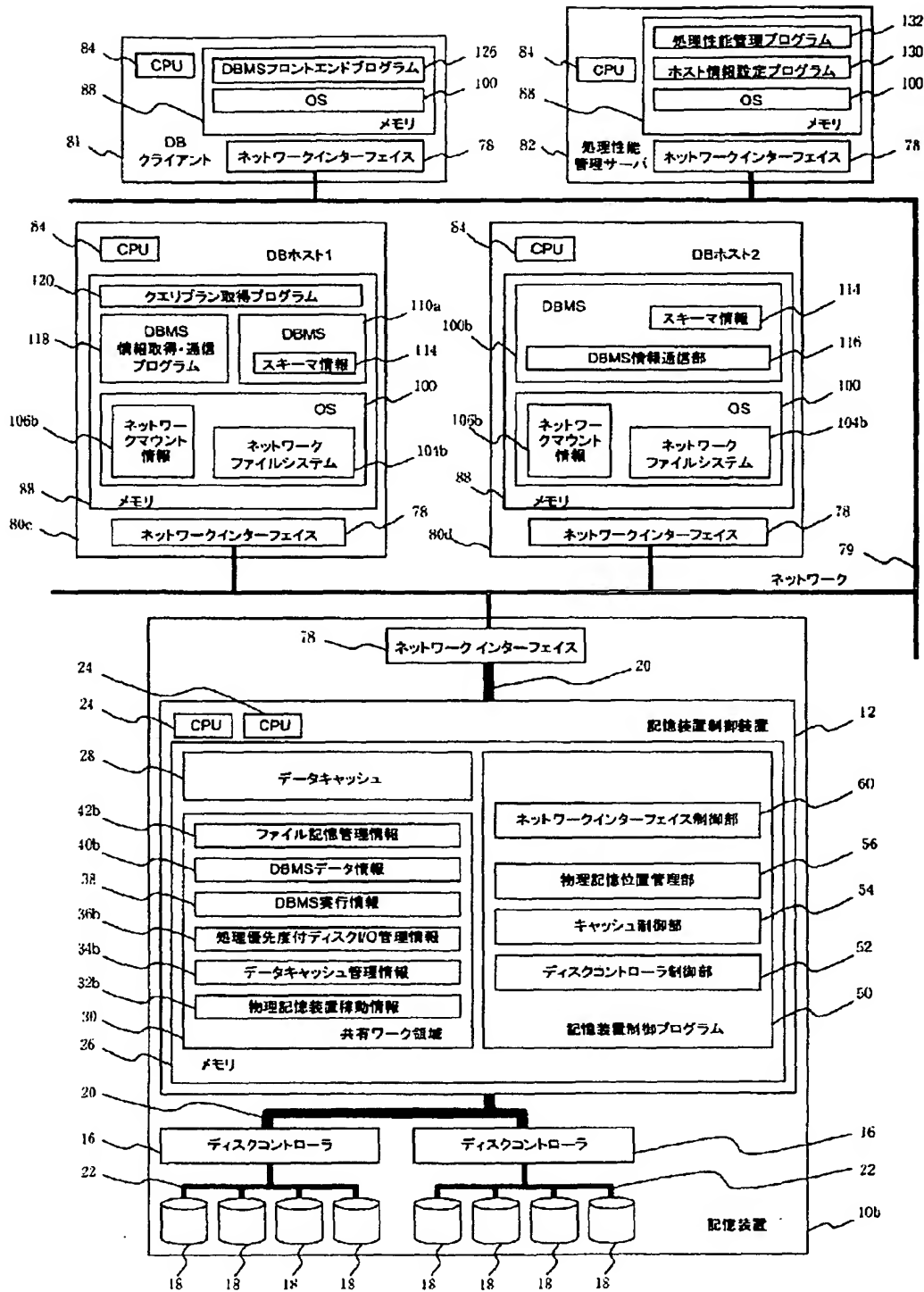


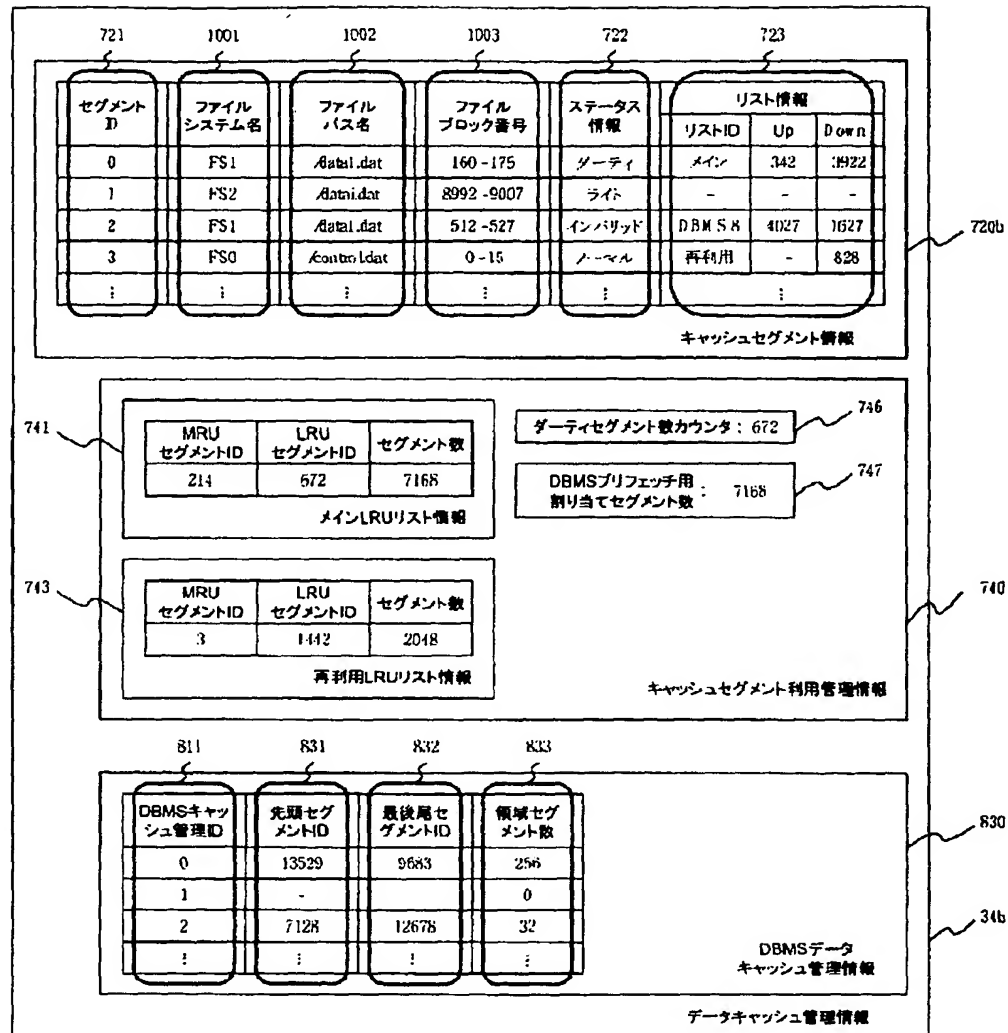
图27

| | | | | | | | |
|-------------|---------------|---------------|---------------|--|---------------|--|------|
| DBMS名 | DBMS1 | DBMS1 | DBMS1 | | DBMS2 | | 631 |
| データ構造名 | T1 | T1 | T2 | | T1 | | 661 |
| データ構造ブロック番号 | 0 - 4999 | 5000 - 9999 | 0 - 4999 | | 0 - 9999 | | 716 |
| ファイルシステム名 | FS1 | FS1 | FS1 | | FS2 | | 1001 |
| ファイルパス名 | /data1.dat | /data1.dat | /data2.dat | | /data1.dat | | 1002 |
| ファイルブロック番号 | 0 - 4999 | 5000 - 9999 | 0 - 4999 | | 0 - 9999 | | 544 |
| 物理記憶領域名 | Pdisk 1 | Pdisk 0 | Pdisk 1 | | Pdisk 2 | | 514 |
| 物理ブロック番号 | 10240 - 15239 | 10240 - 15239 | 15240 - 20239 | | 20180 - 30179 | | 502 |

データ構造物理記憶位置情報

【図28】

図28



フロントページの続き

(72)発明者 喜連川 優
千葉県松戸市二十世紀が丘丸山町17

Fターム(参考) 5B005 JJ12 KK02 LL17 MM04 NN22
5B075 NR03 NR20 PQ32 QS07
5B082 FA12